TOWARDS IMPROVED HYDROLOGIC LAND SURFACE MODELLING: ENHANCED MODEL IDENTIFICATION AND INTEGRATION OF WATER MANAGEMENT

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the School of Environment and Sustainability

University of Saskatchewan

Saskatoon

By

FUAD YASSIN

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Executive Director
School of Environment and Sustainability
University of Saskatchewan
Room 323, Kirk Hall
117 Science Place
Saskatoon, Saskatchewan S7N 5C6
Canada
Phone: (306) 966-1985 Fax: (306) 966-2298
Email: sens.info@usask.ca

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

ABSTRACT

Large-scale hydrological models are essential tools for addressing emerging water security challenges. They enable us to understand and predict changes in water cycle at river-basin, continental, and global scales. This thesis aimed to improve 'land surface models' for large-scale hydrological modelling applications. Specifically, the research contributions were made across four fronts: (1) improving the conventional procedure for parameter identification of hydrological processes by using new sources of remotely-sensed data in addition to streamflow data within a multi-objective optimization and sensitivity analysis framework, (2) developing and integrating an efficient parameterization scheme for the representation of reservoirs into the land surface model for realistic representation of downstream flows, which can further feedback to land surface and atmospheric models, (3) demonstrating how precipitation uncertainty from multiple high-resolution precipitation products influences the performance of a land-surface based hydrological model, and (4) developing an enhanced and comprehensive large-scale hydrologic model for a complex and heavily regulated watershed.

The analyses and results of this thesis illuminated important issues and their solutions in large-scale hydrological modelling. First, the multi-objective optimization and sensitivity analysis approach using multiple state and flux variables and performance criteria enables robust model parameterization and lessens issues around parameter equifinality in the highly-parameterized land surface models. Second, the dynamic parameterization of reservoir operation, based on multiple storage zones and reservoir release targets, improves the simulation of reservoir storage dynamics and downstream release, and subsequently, significantly improves the fidelity of land surface models when modeling managed basins. Third, there is a critical need for a rigorous evaluation of precipitation datasets widely used for forcing land surface models. The datasets investigated here showed considerable discrepancies, bringing their utility for land surface modelling into question. Fourth, effective parameterization and calibration of land surface models is critically important, particularly in large, complex, and highly-regulated basins.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to those who helped me shape my career and significantly improved my PhD research work.

First and foremost, I offer my sincere thanks to my supervisors, Prof. Howard Wheater and Dr. Saman Razavi. Since the beginning of my PhD, their positive impact was huge and they deserve great appreciation for their guidance, optimism, encouragement and support throughout my Ph.D. program. I am so thankful for the opportunity to be part of inspiring project of Changing Cold Region Networks and also very happy to be part of the exciting research environment of the Global Institute for Water Security.

I am grateful to my advisory committee members Prof. Patricia Gober, Dr. Al Pietroniro, Prof. Amin Elshorbagy, and Dr. Karl Lindenschmidt, for their advice and comments that helped to shaping my Ph.D. research work. I would like to extend my deep gratitude to Dr. Karl Lindenschmidt for valuable support as a chair of my Ph.D. committee.

It is my pleasure to thank Daniel Princz and Dr. Gonzalo Sapriza Azuri for all their technical support. I am grateful to all members of the CCRN and GWF large-scale modelling, team particularly to Dr. Al Pietroniro, Dr. Bruce Davison, Dr Mohamed Elshamy, Dr. Samson Mengistu, Dr. Razi Sheikholeslami, and Dr. Amin Haghnegahdar, Dr. Jefferson S. Wong for valuable interactive discussions. I also wish to thank Dr. Kevin Shook and Prof. John Pomeroy for sharing their valuable experience and knowledge.

I gratefully acknowledge the scholarship provided by the School of Environment and Sustainability through the Canada Excellence Research Chair in Water Security, the Saskatchewan Innovation Opportunity Scholarship, and School of Environment and Sustainability returning scholarship.

Above all, my deepest gratitude goes to my mother Rabya Jemal, my brothers and sisters for their encouragement, unconditional love and support.

DEDICATION

I dedicate this thesis to my parents, brothers and sisters

I love you all dearly

IN MEMORY OF MY FATHER, ABDO, AND MY BROTHER, MOHAMMED

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Background and problem statement

Hydrological model development based on physical realism allows understanding of the mechanisms and complex interactions of the underlying physical processes bridging the gap between theory and observations, and extrapolation from available measurements both in time and space into the future where measurements are not possible (*Beven,* 2012). The development of models requires a combination of mathematical approximations of the dominant processes of what are complex systems (*Freeze and Harlan,* 1969*; Abbott et al.,* 1986). For example, hydrological models are built to represent and approximate the governing processes and interactions of the hydrological cycle components such as infiltration, evapotranspiration and runoff, including in cold regions glaciers, snow accumulation, snowmelt, and snow redistribution. Hydrological models with accurate representation and integration of these processes are valuable tools to help manage the world's finite water resources, evaluate the impact of environmental changes, and predict hydrological extreme events such as floods and droughts (*Abbaspour et al.,* 2015*; Abdulla et al.,* 1996*; Beven et al.,* 1984*; Christensen et al.,* 2004*; Dibike and Coulibaly,* 2005*; Döll et al.,* 2003*; Middelkoop et al.,* 2001*; Vörösmarty et al.,* 2000).

Over the last few decades, the development of hydrological models for large-scale application (~10^3-10^6 km^2) has gained momentum because of emerging water security challenges (*Döll et al., 2003; Vörösmarty et al.,* 2000; *Eagleson* 1986). This is because the majority of the practical and scientific challenges to recognize and prepare for environmental change are linked to our capability to understand and predict water cycle changes at large scales, including regional, continental and global scales (*Eagleson* 1986; *Clark et al.,* 2015). Advances in computational power, remote sensing datasets, and climate datasets have been the main contributing factors to the development and growth of several large-scale models that permit model simulations for large geographical domains (*Wood et al.,* 2011; *Bierkens et al.,* 2015; *Melsen et al.,* 2016). Thus far, large-scale hydrological models have proved to be important tools in monitoring water security and evaluating the impact of population increase and environmental changes, e.g. climate change

and land use/cover change, at large scales (a large basin, a continent or the globe) (*Döll et al.,* 2003*; Vörösmarty et al.,* 2000*; Wisser et al.,* 2010).

Catchment hydrological models focus on modelling of natural catchments, development of methods for appropriate process representation (primarily for physically-based models), and developing of datasets and methods for parameter estimation and transferability. Practically, catchment-scale hydrological models are not feasible to be directly used for the large-scale modelling tasks, as they involve more detailed descriptions of processes, requiring large computational resources, large input datasets, and calibration of many parameters (*Sood and Smakthin*. 2015). It is also a question of a "horses for courses" situation whereby small scale processes employed in catchment models are not essentially dominant at larger scales as the process equations are often scale dependent *(Beven,* 1990). Thus, compared to catchment hydrological models, large-scale hydrological models differ in the scale-appropriate simplification of the physically-based hydrologic process representations, and large-scale hydrological models may only involve calibration of few parameters or not be calibrated at all (*Gupta et al.,* 1999; *Davison et al.,* 2016).

There are many large-scale hydrological models, which broadly fall into two categories, namely: Global Hydrological Models (GHMs) and Land Surface Models (LSMs) (*Döll et al.,* 2016; *Gudmundsson et al.,* 2012). GHMs focus on water balance-based simulations of global hydrology based on conceptual model approaches dominantly derived from catchment-scale models. Moreover, they aim to improve scale-appropriate process representations mainly for water management purposes (water demand based on population, livestock, irrigation schemes and industrial water use) in order to address water security challenges and vulnerabilities to climate change and growing human water use at a global scale (*Döll et al.,* 2003*; Vörösmarty et al.,* 2010). Examples of models in this group include the Water-Global Analysis and Prognosis model (WaterGAP) (*Doll et al.,* 2003; Alcamo *et al.,* 2003), PCRaster GLOBal Utrecht University (PCR-GLOBWB) (*van Beek and Bierkens* 2008), Water Balance Model- Water Transport Model (WBM-WTM) (*Vörösmarty et al.,* 1998), Lund-Postdam-Jena managed Land model (LPJmL) (*Bondeau*

et al., 2007), and Global Water Availability Assessment model (GWAVA) (*Meigh et al.,* 1999), and the Variable Infiltration Capacity model (VIC) (*Liang et al.,* 1994). GHMs are usually run at a coarse spatial resolution (between 0.5° and 1° dictated by the resolution of forcing data) and sub-annual temporal resolution (mostly monthly).

LSMs were originally built to provide lower boundary conditions to climate models by simulating the water and energy balances on the land surface. They control the complex exchanges of water, energy and, sometimes, carbon between the land, the atmosphere, and the ocean. The well-established recognition of the importance of the land surface processes for climate models transformed LSMs from bucket hydrology (1st generation LSMs) towards improved process-based hydrology (3rd generation LSMs) (*Pitman,* 2003). The significant advances in LSMs have motivated their use beyond their original purpose, now to simulate large-scale hydrology by running them in an offline-mode (*Clark et al.,* 2015*; Davison et al.,* 2016). Models in this group include the Community Land Model (CLM) (*Oleson et al.,* 2013), Modélisation Environmentale communautaire - Surface Hydrology (MESH) (*Pietroniro et al.,* 2007), Organizing Carbon and Hydrology in Dynamic Ecosystems model (ORCHIDEE) (*d'Orgeval et al.,* 2008), NOAH-MP (*Yang et al.,* 2011), Joint UK Land Environment Simulator JULES (*Best et al.,* 2011), and Minimal Advanced Treatments of Surface Interaction and Runoff (MATSIRO) (*Takata et al.,* 2003). The grid size usually ranges from 0.5° and 1°, but higher resolutions can be used for regional scale applications. LSMs require climate forcing at high temporal resolution (typically $\leq$ 3hrs), and tend to require more climate forcing variables than catchment hydrological models to conduct the energy and water balances. LSMs often do not include routing and model parameter calibration is not routinely conducted.

To improve the utility of LSMs for large-scale hydrological modelling purposes, there are relevant outstanding issues that are yet to be addressed for better representation of the dominant hydrological processes and human interventions (*Clark et al.,* 2015*; Davison et al.,* 2016*; Nazemi and Wheater*, 2015a *and* 2015b*; Gupta et al.,* 2008*; Wada et al.,* 2017*; Pokhrel et al.,* 2016; *Archfield et al.,* 2015). The review of *Davison et al.* (2016) focused on describing the missing

lateral processes representations in most LSMs including river routing, blowing snow, reservoir routing, wetland dynamics, overland flow, interflow, groundwater flow, and flow diversion. The study compared nine LSMs to demonstrate which models include which lateral processes. Additionally, *Davison et al.* (2016) showed the benefit of LSM parameter calibration for simulation of streamflow and latent heat flux on an experimental watershed. *Clark et al.* (2015) provided summaries and comparisons on how LSMs represent different hydrological processes with more emphasis on storage and transmission processes including infiltration, water movement in soil, root water uptake, groundwater dynamics, stream-aquifer interactions, and channel/floodplain routing. *Clark et al.* (2015) provided further guidance on the key opportunities to improve dominant hydrological processes in LSMs. On the subject of adequate representation of human interventions in LSMs, *Nazemi and Wheater* (2015a and 2015b), *Wada et al.* (2017), and *Pokhrel et al.* (2016) described the challenges and systematic approaches needed towards full integration of these in large-scale models.

This thesis focuses on the second group of large-scale models (LSMs). The choice to proceed with LSMs is not because LSMs are better than GHMs, but rather the need for improving the utility of LSMs for large-scale hydrologic modelling purposes. Development of LSMs can potentially facilitate further improvement in climate models because LSMs are an integrating interface between climate and hydrological modelers, providing a common framework to converge scientific model development from both modelling communities. Thus, any improvement in hydrologic processes and water management representation made in LSMs can be possibly integrated into climate models in a coupled mode.

In this thesis, building upon previous studies, some key opportunities to improve LSMs for large-scale hydrological modelling are explored. In particular, this thesis was designed to address the following three issues:

- The proper values for LSM parameters are typically chosen rather arbitrarily, and as a common approach parameters values are pre-specified by referring to look-up tables

(*Verseghy*, 2000; *Mendoza et al.,* 2015). The parameters are rarely calibrated largely due to the difficulties of managing the large number of inter-related parameters that are used to describe the underlying processes of vegetation, soil and snow components. Even though more complex models provide more flexibility for better representation, it is challenging to identify optimal parameterization to complex models because the equifinality problem increases with complexity. The common approach of using a-priori parameter values usually produces a sub-optimal model simulation of states and fluxes. The importance of proper parameter estimation on improving model outputs has been demonstrated in some studies, mostly using streamflow observations to obtain optimal parameter values (*Gupta et al.,* 1999; *Nasonova et al.,* 2009). However, calibration using streamflow alone is not sufficient, since LSMs make use of a large number of model parameters, which if not properly constrained, may lead to poor model performance for other model states and fluxes. For enhanced parametrization, it is essential to take into account observations of other variables such as evapotranspiration, soil moisture, and terrestrial water storage anomaly from satellites (*Livneh and Lettenmaier*, 2012; *Lo, et al.,* 2010; *Werth and Güntner*, 2010).

- The global, regional, and local water cycles have been modified by human interventions through damming of rivers, diversion of flow, and abstraction of water from surface and groundwater resources to satisfy consumptive and non-consumptive water demands *(Nazemi and Wheater*, 2015a, 2015b). Failing to represent human interventions within the modelling framework of LSMs (and large-scale models in general) limits their performance in the many highly regulated basins across the globe. It also constrains their applicability in investigating the future potential states of watershed systems under different scenarios (e.g., scenarios of climate, land use, or reservoir regulation changes). Adequate representation of human interventions remains as a common challenge to LSMs, and the existing methods still require further refinement. For example, the representation of reservoir operations in LSMs is overly simplified (*Meigh et al.,* 1999; *Döll et al.,* 2003;

Hanasaki et al., 2006) and does not accurately reproduce reservoir storage and downstream release.

- There has been limited research for the comprehensive evaluation of LSMs in complex, highly regulated basins under the constraints of significant uncertainty in precipitation and parameters. Most LSMs use energy-based approaches that require high spatiotemporal resolution climate forcing data. Due to limitation of observed high resolution climate forcing data, LSMs are often driven by gridded reanalysis climate forcing data derived from numerical weather prediction model outputs (*Sheffield et al.,* 2006; *Weedon et al.,* 2014; *Côté et al.,* 1998; *Mesinger et al.,* 2006). Although reanalysis and interpolated data products provide better coverage of the space-time field of the meteorological conditions, they contain significant errors that vary from one data product to another. The accuracy of each climate forcing product across different regions can have substantial impacts on the simulation of streamflow (*Eum et al.,* 2014). The errors in a climate forcing product when combined with model parameter calibration can mask major model deficiencies, as different components, such as model structure, model parameters and forcing data can compensate for the shortcomings and error of each other, so that improvements to streamflow simulation calibration may be at the cost of degrading other model outputs. Thus, it is important to 1) identify the sources of uncertainty, and quantify their impact on model simulation, and 2) evaluate the model performance on other flux and storage outputs before and after calibration, to ensure the effective use of the LSMs and to increase the reliability of model predictions, especially for large-scale basins that are increasingly influenced by human activities.

1.2 Research objectives and hypotheses

The overall objective of this study is to improve the capability of the LSM framework in reproducing historical hydrological output variables and in projecting the impact of future environmental change on hydrology and water resources. To demonstrate and address these thesis

objectives, a Canadian LSM framework called MESH was used as a platform. MESH has been used extensively for Canadian (cold-region) watershed research studies (*Pietroniro et al.,* 2007; *Davison et al.,* 2006; *Mekonnen et al.,* 2014; *Haghnegahdar et al.,* 2014).

The rationale or hypotheses of the thesis include:

1) Multi-objective optimization based on model storage and flux outputs is necessary for enhanced hydrological model identification under high-dimensional parametrization problem,
2) A reservoir operations model with a dynamic characterization of reservoir storage zones and releases is essential to improve the reservoir model simulations of storages and releases, and hence to provide a better representation of reservoirs in large-scale hydrologic models compared to existing approaches, and
3) Comprehensive evaluation of LSMs that includes and addresses precipitation uncertainty, model parameter calibration, and multiple types of observations for model validations is much needed to reveal LSM utility in large, regulated basins with significant human intervention.

This thesis aims to contribute to the following three interrelated research objectives that will contribute to advancing LSM utilization for large-scale hydrological modelling.

1.2.1 Research objective I

The first objective aims to improve the parameter estimation of LSMs with a large number of parameters describing vegetation, soil, snow, and routing processes. Multi-criteria sensitivity analysis and optimization approaches were used to enhance the identification of model parameters using streamflow and GRACE total water storage (GRACE-TWS) anomaly observations simultaneously.

The novelty and significance of this research lies in the developed multi-criteria sensitivity and optimization framework to utilize the GRACE satellite storage data as a supplement to streamflows

for model calibration. The developed framework is general and allows the utilization of any supplementary storage and flux data to reduce the heavy reliance on the conventional streamflow-based model calibration, and to develop a high-fidelity model, particularly when optimizing a large number of parameters.

Under this research component, methods were designed to address the following research questions: Question 1: *How does the sensitivity of model parameters change when the response function is switched from metrics on streamflow to metrics on storage variables?* Question 2: *How does the model performance improve as a result of calibration to GRACE-TWS in addition to streamflow?* Question 3: *How does parameter identifiability improve as a result of calibration to GRACE-TWS in addition to streamflow?*

1.2.2 Research objective II

The objective of this research component is to develop an improved reservoir operation model with a flexible level of complexity, which allows adequate capturing of the regulation introduced along a river system by dams and reservoirs. Subsequently, the developed reservoir operation model is integrated into the LSM to evaluate its performance at the LSM temporal resolution, in combination with other LSM processes.

The novelty and significance of this section are twofold: the primary objective is to configure the reservoir model to have multiple storage zones that can vary dynamically at selected time intervals; and the secondary objective is to develop a generic and efficient parameterization approach that can deduce reservoir operation rules from observed concurrent series of reservoir inflow, storage, and outflow data. The reservoir model significantly improves the representation of reservoir storage dynamics and downstream releases, compared to existing reservoir models and provides a better representation of reservoirs in large-scale hydrologic models.

To address our objectives and test our hypothesis, the experiment of this study is designed to address the following research questions. Question 1: *To what extent does the developed reservoir model improve model simulation compared to ignoring reservoirs in the model and compared to*

other widely used reservoir models in LSMs? Question 2: *How sensitive is the developed reservoir model release and storage performance to initial reservoir storage value and inflow bias?* Question 3: *How does the parameterization of the developed reservoir model through calibration compare against no calibration reservoir parameterization?* Question 4: *How well does the integration of the developed reservoir model into the LSM help capture observed reservoir storage and release?*

1.2.3 Research objective III

The third objective is to conduct a detailed analysis and evaluation of a physically-based Canadian LSM for a highly managed, large-scale basin, using state-of-the art calibration strategies and multiple data sources to enable quantification of modelling uncertainty. The objective under this component has multiple elements: (1) to evaluate the quality of gridded precipitation datasets in terms of how well they reproduce observations of multiple streamflow gauges when used to drive LSM, (2) to improve the H-LSM parameterization using a state-of-the-art, computationally-efficient calibration approach, and evaluate the effectiveness of parameter transferability through validation in time and space, using independent multiple streamflow gauges not used in calibration, (3) to test the model performance using multiple sources of observational information on model storage and output fluxes, and to ensure that the optimal parameters obtained are as realistic as possible (giving the "right answers for the right reasons") without error compensation across multiple outputs.

The primary novelty and significance of this section are the examination of precipitation uncertainty based on how well they simulate streamflow observed at multiple locations across the basin. Second, a state-of-the-art multi-criteria calibration approach was applied, using various observational information including streamflow, storage and fluxes for calibration and validation. Furthermore from a technical point of view, and for the purpose of comprehensive benchmarking of model performance, this section includes assembling/introducing an irrigation algorithm, the

proposed reservoir operation model, diversions, and gridded sub-module control, irrigated land fraction, and gridded soil texture to the model setup of the Canadian LSM.

Overall, this component aims to address: Question 1: *How do the error characteristics of the widely-used high-resolution datasets vary from one dataset to another over a large-scale basin?* Question 2: *How does precipitation error affect the reliability of model streamflow simulations?* Question 3: *To what extent does model calibration improves the default parameterization performance?* Question 4: *Can calibrated parameters on subbasins be used as a global parameter set for the whole basin?* Question 5: *Does the complex configuration of MESH provide adequate simulation of different basin storage and flux outputs?*

1.3 Thesis structure

The thesis is organized based on the 'dissertation by manuscript' style. The thesis, starts with an introduction (this chapter) followed by three chapters (Chapters 2 to 4) that present three manuscripts addressing the three research objectives/components listed above. Therefore, each chapter from chapters 2 to 4 present its own research objectives and questions, which will address components of the overarching research objective of this thesis. Each manuscript chapter (Chapters 2 to 4) also contains a comprehensive literature review demonstrating the novelty and significance of the research in addressing the questions. In addition, each chapter presents its own methodologies and results in addressing the research questions. Finally, this thesis ends with Chapter 5, which provides a summary of the research work and provides directions for future work.

CHAPTER 2

ENHANCED IDENTIFICATION OF A HYDROLOGIC MODEL USING STREAMFLOW AND SATELLITE WATER STORAGE DATA: A MULTI-CRITERIA SENSITIVITY ANALYSIS AND OPTIMIZATION APPROACH

This chapter is based on the following published article.

Yassin F, Razavi S, Wheater H, Sapriza-Azuri G, Davison B, Pietroniro A. 2017. Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. Hydrological Processes 31 (19): 3320–3333 DOI: 10.1002/hyp.11267 https://onlinelibrary.wiley.com/doi/full/10.1002/hyp.11267

Author contributions: FY developed the method and experiments, wrote the necessary computer codes, coupled the model with a multi-objective optimization scheme, and performed all numerical experiments. FY, SR and HW contributed to the interpretation of the results, structuring and formulation of the paper. FY wrote the paper with contributions from SR and HW. All co-authors contributed to editing of the paper.

Synopsis

Hydrologic model development and calibration has continued in most cases to focus only on accurately reproducing streamflows. However, complex models, for example so-called physically-based models, possess large degrees of freedom that, if not constrained properly, may lead to poor model performance when used for prediction. We argue that constraining a model to represent streamflow, which is an integrated resultant of many factors across the watershed, is necessary but by no means sufficient to develop a high-fidelity model. To address this problem, we develop a framework to utilize the Gravity Recovery and Climate Experiment's (GRACE's) total water storage anomaly data as a supplement to streamflows for model calibration, in a multi-objective setting. The VARS method (Variogram Analysis of Response Surfaces) for global sensitivity analysis is used to understand the model behavior with respect to streamflow and GRACE data, and multi-objective optimization method is applied for model calibration. Two sub-basins (Battle and Vermilion) of the Saskatchewan River Basin in Western Canada are used as a case study. Results show that the developed framework is superior to the conventional approach of calibration only to streamflows, even when multiple streamflow-based error functions are simultaneously minimized. It is shown that a range of (possibly false) system trajectories in state variable space can lead to similar (acceptable) model responses. This observation has significant implications for land-surface and hydrologic model development, and, if not addressed properly, may undermine the credibility of the model in prediction. The framework effectively constrains the model behavior (by constraining posterior parameter space) and results in more credible representation of hydrology across the watershed.

2.1 Introduction

Hydrologic models are becoming more complex, as we aim to rigorously simulate our growing understanding of the underlying natural processes. More complexity, however, does not necessarily come with improved fidelity of a model to the real-world system. One problem is that more complex dynamical systems models typically have more 'degrees of freedom' that may result

in a greater number of independent trajectories that a dynamical system can evolve through time without violating imposed constraints (*Layek,* 2015). The degrees of freedom of such models are related to their large state space that embeds various state variables across the spatial domain, in addition to their many parameters needed to enable the representation of many processes. The possession of such large degrees of freedom or flexibility can be an asset of the current generation of hydrologic models as it enables reproduction of a wide variety of possible hydrologic scenarios; this flexibility, however, if not constrained properly, can lead to poor model performance, particularly in validation.

Hydrologic models treat the catchment as a dynamical forcing-state-response system. Calibration is an essential building block for the identification of the majority of such models, where model parameters are tuned to maximize the 'fidelity' of the model to the underlying real-world system. The effective characterization of 'fidelity' itself, however, is often nontrivial and has been typically limited to metrics that assess the goodness of fit of these models to streamflow data collected for a limited number of years. Calibration to streamflows is appealing because:

1. Streamflow is an integrated response of a watershed and therefore calibration to streamflows controls many model processes,
2. Streamflow is measured relatively easily and the extent of uncertainty in measured streamflows is typically known (e.g., uncertainty in rating curves), and
3. Streamflow is often the most important hydrologic variable and is the basis for water management.

The 'conventional approach' to the assessment of model fidelity that only relies on streamflows has been widely used with various methods for watershed systems analysis, including the use of single- and multi-criteria optimization (*Duan et al.,* 1992; *Gupta et al.,* 1998; *Razavi and Tolson,* 2013; *Razavi et al.,* 2010), Bayesian inference and uncertainty analysis (*Beven and Binley,* 1992; *Liu and Gupta*, 2007; *Vrugt et al.,* 2003), multiple-hypothesis approaches (*Clark et al.,* 2008), and sensitivity analysis (*Razavi and Gupta,* 2015). In addition to model fidelity metrics

that are based on residuals, there have been studies to characterize and reproduce hydrologic signatures (e.g., flow duration curves) of watersheds embedded in streamflows through diagnostic approaches (*Wagener et al.*, 2007; *Gupta et al.*, 2008) or multi-objective calibration (*Pechlivanidis et al.*, 2014).

There have been studies to improve the characterization of model fidelity by incorporating other types of observations for different model state and response variables, in addition to streamflows. *Kuczera and Mroczkowski* (1998) supplemented streamflow data with groundwater level and water salinity data to improve parameter identifiability of a conceptual hydrologic model in a Bayesian inference framework. *Madsen* (2003) utilized streamflow and groundwater level data in a weighted multi-objective optimization for parameter calibration of the MIKE-SHE distributed hydrologic model. *Bekele and Nicklow* (2007) calibrated the SWAT model to streamflow and sediment data by an evolutionary multi-objective optimization algorithm. As a supplement to quantitative observations in the type of analyses mentioned above, *Seibert and McDonnell* (2002) raised the significance and utility of qualitative information on catchment behavior (i.e., knowledge from experimentalists) for model calibration. They formulated a multi-criteria calibration approach that formally utilizes such so-called "soft data" (e.g., new water ratio at flow peaks) in conjunction with the commonly used streamflow and groundwater level data.

Recent research has begun to explore the potential of different remotely-sensed data products to constrain model performance, typically at the continental scale (*Werth et al.*, 2009; *Werth and Güntner*, 2010; *Lo et al.*, 2010; *Livneh and Lettenmaier*, 2012; *Silvestro et al.*, 2015; *Rakovec et al.*, 2016; *Qiao et al.*, 2013). A major opportunity is the Gravity Recovery and Climate Experiment (GRACE), which is the only satellite mission currently capable of detecting changes in total water storage (TWS), under any condition, by making detailed measurements of Earth's gravity field anomalies (*Zaitchik et al.*, 2008). Given the availability of GRACE data for more than 10 years and the planned GRACE-FO (Follow-On) mission to be launched in 2017, it provides a unique dataset to improve our understanding of change in TWS in large spatiotemporal domains across large watershed systems. The challenges with use of GRACE data are, however, their coarse

spatiotemporal resolution, and the varying uncertainty due to different sources, such as uncertainties in removing ocean, atmospheric, and isostatic rebound signals, gravity measurement noise, and uncertainties in filtering and scaling approaches for different geographical regions (*Seo et al.*, 2006; *Landerer and Swenson*, 2012; *Lambert et al.*, 2013). *Werth et al.* (2009) and *Werth and Güntner* (2010) demonstrated the value of the GRACE data for improved monthly continental water storage simulations across a number of major basins worldwide, by calibrating the WaterGAP in a bi-objective setting to monthly river discharge and GRACE data. *Lo et al.* (2010) combined GRACE and baseflow data in a weighted error function for Monte-Carlo-based calibration of the Community Land Model (CLM) to improve shallow groundwater level simulations in Illinois, United States. *Livneh and Lettenmaier* (2012) evaluated multiple types of remotely-sensed data, including the GRACE data, to supplement streamflow observations in the calibration of the Unified Land Model, on numerous basins of the continental United States.

Despite some success in improving the model performance, they concluded the uncertainties in remote-sensing data to be a limiting factor in their utility for parameter estimation. *Silvestro et al.* (2015), however, showed the usefulness of remotely-sensed land-surface temperature and surface soil moisture data (combined with streamflow data via a weighted error function) for hydrologic model parameter identification in Italy. The utility of the GRACE data is continually increasing because of the increase in length of the collected data and the frequent refinements made to improve the accuracy of its processed TWS data.

2.2 Objectives and scope

This research and conducted experiments were designed to address the following questions:

- Question 1: How does sensitivity to different model parameters change when engaging state variables in calibration?
- Question 2: How does the model performance improve as a result of calibration to GRACE-TWS in addition to streamflows?

- Question 3: How does parameter identifiability improve as a result of calibration to GRACE-TWS in addition to streamflows?

First, for Question 1, we conducted a global sensitivity analysis (GSA) to independently characterize the sensitivity of the model state variables and responses to a range of model parameters. These experiments were designed to identify the different dominant model parameters that control the model capability to reproduce observed streamflow and/or GRACE-TWS data. For Questions 2 and 3, we developed two multi-objective optimization (MOO) problems, to understand the value added (information extracted) by constraining model state variables from a fully multiple-criteria perspective. This design of the two MOO problems provided two unique features:

I. Clearer characterization of information gained via GRACE-TWS: In the objective function space, the Pareto-solutions to the optimization problem 1 (i.e., with only streamflows) are a sub-set of the Pareto-solutions to the optimization problem 2 (i.e., with GRACE-TWS and streamflows) (related to Question 2 above), while this is not necessarily the case in the parameter (decision variable) space (related to Question 3 above). The new dimension (the error metric on TWS) introduced by the optimization problem 2 could effectively demonstrate the value of the new information in improving the model performance and parameter identification.

II. Better treatment of uncertainty in GRACE-TWS data: uncertainty in remotely-sensed and streamflow data can be of different forms and magnitudes, and may not be treated equally. The multiple-criteria nature of the optimization problem 2 allows that GRACE-TWS data (and the associated uncertainty) to be treated (and relied on) flexibly to whatever extent needed. In other words, this design enables the user to understand and navigate the Pareto front between the error metric on TWS and the ones on streamflow, and stop where appropriate.

The MESH Land-Surface and Hydrology modelling system (*Pietroniro et al.*, 2007) is used in this study. Two prairie basins of the Saskatchewan River Basin (SaskRB) in western Canada are used as the case study (see **Fig.1a**). The SaskRB is a Regional Hydroclimate Project of the World Climate Research Programme's GEWEX project (*Wheater*, 2013), and experiences one of the most extreme and variable climates in the world and embodies a set of critical challenges for water security, which are of particular importance to western Canada and relevant globally (*Razavi et al.*, 2015; *Wheater and Gober*, 2013, 2015).

The organization of the remainder of the paper is as follows. Section 2.3 presents our approach to model calibration, the tools used, and the developed experiments. Section 2.4 provides background information about the case study and data used for model calibration and validation. Section 2.5 presents the results of the experiments, including parameter sensitivity analysis, and optimization. The paper ends with conclusions in Section 2.6.

2.3 Methodology

2.3.1 Design of experiments

The GSA experiment using the VARS method (see section 2.3.2 for details on VARS) was conducted to identify the dominant model parameters that control the model capability to reproduce observed streamflows and GRACE TWS anomalies. To simulate TWS anomalies in the model, we accounted for the water storage in surface ponding, snow, frozen and liquid soil water content of all layers, and liquid and snow storage on the canopy. For the purpose of the sensitivity analysis, Nash-Sutcliffe Efficiency was used to measure the goodness-of-fit to daily streamflows (called NSE (FL) hereafter) and GRACE-based monthly TWS (called NSE (TWS) hereafter).

The two MOO experiments for model calibration were carried out using the BORG algorithm (*Hadka and Reed*, 2013). The first MOO experiment was based on the classic approach (called 'approach 1' hereafter), where the following three complementary objective functions on streamflows were defined and optimized: NSE (FL) that emphasizes high flows, NSE on logarithm

of streamflows (NSE (logFL)) that emphasizes low flows, and percent bias (PBIAS (FL)) that measures errors in simulation of the total volume of streamflow. Approach 1, which is a three-objective optimization problem, may be deemed the current standard practice for calibration to streamflows. The second MOO experiment followed our new approach (called 'approach 2' hereafter), where, in addition to the three objective functions defined above, we defined the fourth objective function to be NSE (TWS). This objective function extracts the modelled TWS anomaly over time across the basin and compares it against GRACE data.

Notably, the three-objective optimization problem of approach 1 is a sub-problem of the four-objective optimization problem of approach 2. This means that, theoretically, the Pareto front of approach 1 in the objective function space is most likely a part of the Pareto front of approach 2. In the parameter space, however, the Pareto parameter sets identified by the two approaches may not overlap, unless there is only one region of attraction in the multi-objective problem response surface (this is very unlikely given the multi-modality of such response surfaces). Approach 2 can be seen as an extension to approach 1 enhanced by the addition of a new constraint that is handled as a new objective function (like a penalty function). When this constraint is relaxed, approach 2 reduces to approach 1; whereas by tightening the constraint, the calibration problem tends to model parameters that are optimal with respect to all the metrics (including the new one).

2.3.2 Global sensitivity analysis

Global sensitivity analysis is "a systems theoretic approach to characterizing the overall (average) sensitivity of one or more model responses across the factor space, by attributing the variability of those responses to different controlling (but uncertain) factors (e.g., model parameters, forcings, and boundary and initial conditions)" (*Razavi and Gupta*, 2016a). Among several existing global sensitivity analysis approaches, the variance-based (e.g., Sobol (*Sobol'*, 2001) and derivative-based (e.g., Morris (*Morris* 1991)) approaches are widely used. In variance-based sensitivity analysis, the sensitivity of a parameter is determined based on its contribution to the total variance of the

response surface. The total variance of the response surface is a combination of effects resulting from each parameter's direct contribution (referred to as "first-order effect") and interaction effects between parameters (higher-order effects). "Total-order effect" of a parameter, which is the combined effect of its first-order effect and all interaction effects of any order with the other parameters, is a common metric to characterize global sensitivity. More on theoretical bases of the variance-based approach can be found in *Sobol'.* (2001), *Saltelli et al.* (2008), and *Razavi and Gupta* (2015). The derivative based approach (e.g., *Morris*, 1991), however, has its roots in local sensitivity analysis where the local effect of parameter perturbation on a model response (referred to as "elementary effect") is evaluated across the parameter space, and then the local effects are averaged to generate a global effect to characterize global parameter sensitivity. The variance- and derivative-based approaches are quite different, and therefore, can lead to significantly different sensitivity results.

In this study, we utilized the recently proposed variogram-based framework, called VARS (Variogram Analysis of Response Surfaces) that bridges the variance- and derivative-based approaches (*Razavi and Gupta*, 2016a and b). VARS provides a comprehensive spectrum of information about the underlying sensitivities of a response surface to its factors while reducing to well-known and commonly used approaches to 'global' sensitivity analysis such variance- and derivative-based approaches as limiting cases (*Razavi and Gupta* 2016a and b). The VARS framework is unique in that it characterizes important sensitivity-related properties of response surfaces including local sensitivities and their global distribution, the global distribution of model responses, and the structural organization of the response surface (*Razavi and Gupta* 2016a; *Sheikholeslami et al.*, 2017).

A brief description follows of how VARS was used to identify the sensitivity of parameters on the response surface. Assume the model response surface $y(\boldsymbol{X}) = \boldsymbol{f}(x_1, x_2, ..., x_n)$, where the vector $\boldsymbol{X} = \{x_1, x_2, ..., x_n\}$ represents the n model parameters that are used to determine the response surface. For a given two points in the parameter space that is $\boldsymbol{X}^a$ and $\boldsymbol{X}^b$ with separation distance $\boldsymbol{s} = \boldsymbol{X}^a - \boldsymbol{X}^b$, the variogram measures the dissimilarity of the response surface with

respect to $\boldsymbol{s}$, as shown in Eq. (1), hence accounting for all combinations of the response surface dissimilarity at a distance of $\boldsymbol{s}$, which is a general indicator of the response surface sensitivity at a scale of $\boldsymbol{s}$ perturbation. Eq. (2) shows the numerical formulation used to identify the variogram in the parameter i direction.

$$\gamma(\boldsymbol{s}) = \frac{1}{2} E\left[\left(y(\boldsymbol{X} + \boldsymbol{s}) - y(\boldsymbol{X})\right)^2\right] \quad (1)$$

$$\gamma(\boldsymbol{s_i}) = \frac{1}{|N(\boldsymbol{s})|} \sum_{(i,j)\in N(\boldsymbol{s})} \left(y(\boldsymbol{X}^a) - y(\boldsymbol{X}^b)\right)^2 \quad (2)$$

$$\Gamma(S_i) = \int^{S_i} \gamma(s_i)\, ds_i \quad (3)$$

Where E represent the expected value, $N(\boldsymbol{s})$ is the set of all possible combination of points in the parameter space in which $\boldsymbol{s} = \boldsymbol{X}^a - \boldsymbol{X}^b$

VARS is enabled with a star-based sampling strategy or STAR (collectively called STAR-VARS) developed in Razavi and Gupta (2016b). STAR is a structured sampling scheme that is partially randomized via Latin hypercube sampling which facilitate the regularity of parameter sampling and variogram estimation across each parameter dimension (see Razavi and Gupta 2016b). To generate a comprehensive metric for global sensitivity, VARS integrates variograms over the range of scale S_i (Eq. 3) (e.g., 10%, 30%, or 50%), while generating variance-based total-order effects and elementary effects as by-products. The integration across scales is referred to as IVARS (Integrated Variograms Across a Range of Scales) and the different level of scale integration provides sensitivity information at different scales. In this study, we used IVARS50, which refers to the integration of a variogram over 50% of the parameter range, as recommended by *Razavi and Gupta* (2016a). VARS and its IVARS metrics are the only available means that formally account for "parameter perturbation scale" that can be defined as a prescribed size of the sensitivity related neighborhood around any point in the parameter space (*Haghnegahdar and Razavi*, 2017).

We conducted one trial of STAR-VARS with the following sampling setting: number of stars = 20, resolution=0.1, and bootstrap size=1000, resulting in 22,550 model runs. Detail of the VARS framework and sampling strategy is available in *Razavi and Gupta* (2016a and b).

2.3.3 Multi-objective optimization

The BORG multi-objective evolutionary algorithm is an effective and efficient algorithm designed for complex, many-objective, multimodal optimization problems (*Hadka and Reed*, 2013, 2014). BORG unifies several state-of-the-art operators and strategies for multi-objective optimization and represents a class of algorithms whose operators are adaptively selected and applied in the course of optimization based on their utilities. BORG has proved to be very efficient, particularly for many-objective (four or more objective functions) optimization problems (*Hadka and Reed*, 2013). Features of Borg include ε*-dominance* that guarantees the diversity and convergence of the Pareto front, ϵ*-progress* that detects search stagnation at local optima to revive the search by randomized restarts, and *auto-adaptive multi-operator recombination* that favors and selects operators that work best on a given problem. For BORG parameters, we followed the recommendations of *Hadka and Reed* (2013, 2014). In particular, initial population size=100, a population to archive ratio γ=4 and a selection ratio of τ=0.02. We used an ε value of 0.001 for all objective functions, which means that a change in any of the objective functions that is greater than or equal to 0.001 is considered as improvement or progress. To account for stochasticity of the optimization process, we ran five replicates (trials) of each optimization experiment with different random seeds. The final Pareto front presented for each experiment was obtained by merging the Pareto fronts of the associated five replicates (non-dominated solutions across the five replicates).

2.4 Case study and data

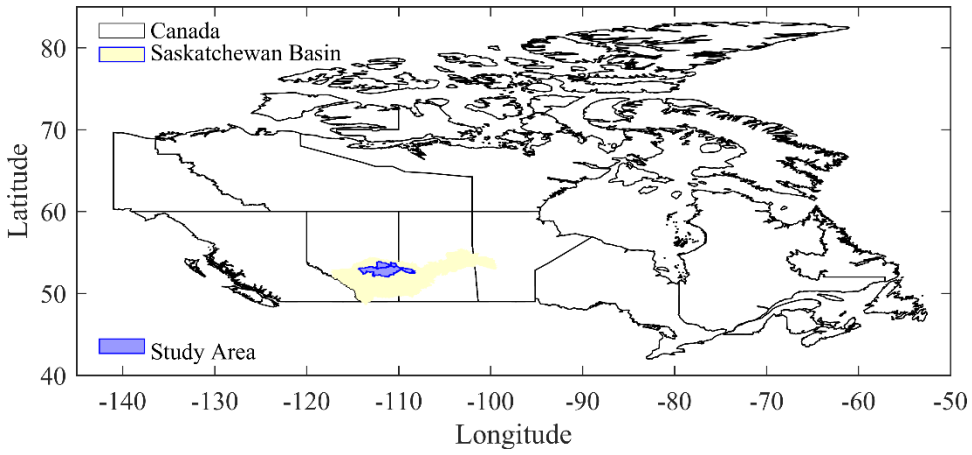
2.4.1 Study area

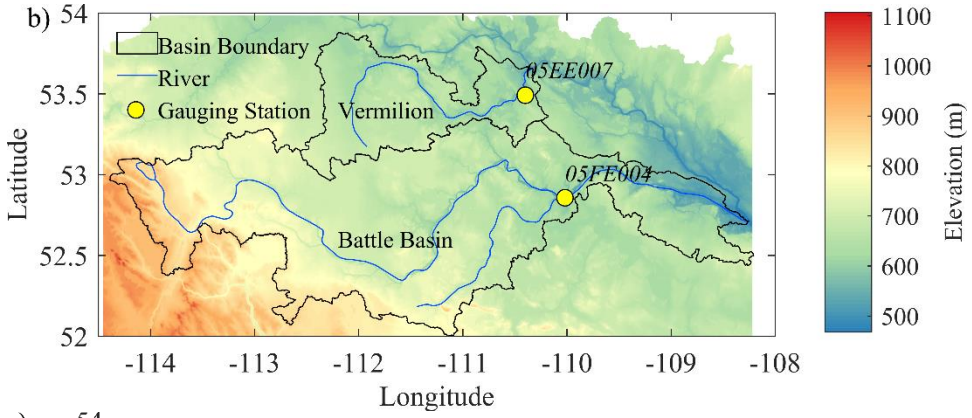
A prairie portion of the SaskRB, located in Western Canada, was selected as a case study. This area consists of the two adjacent Battle and Vermilion River Basins (**Fig. 2.1a** and **Fig. 2.1b**).

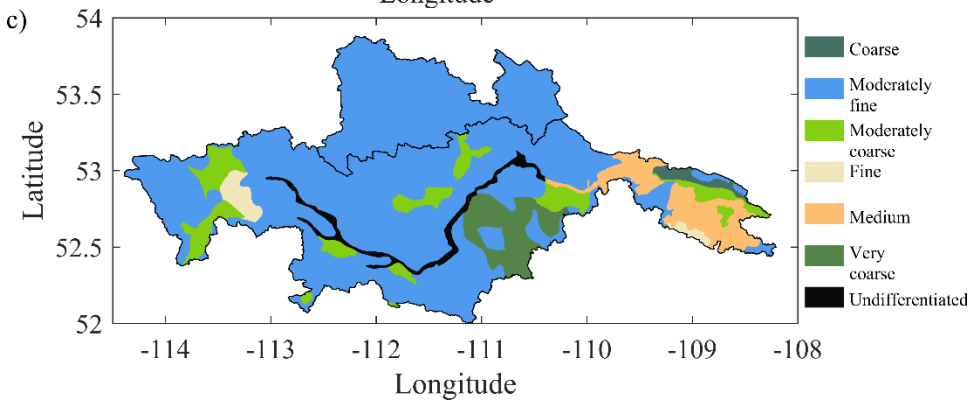
These two basins selected because of the minimal presence of water management compared to other basins within SaskRB. The Battle Basin with an area of 25,100 km^2 (upstream of gauge 05FE004) is used for model calibration and validation in time and, the Vermilion basin with an area of 7,260 km^2 (upstream of gauge 05EE007) was used for model validation in space and time. A 10-year period of 2002 to 2011 was used in the analysis. This period was split into spin-up (2002), calibration (2003-2008), and validation (2009-2011) periods. The majority of surface flows in these basins originate from snow melt (during a relatively short spring period). The annual precipitation in this region varies between 300-500 mm.

The topographic data were obtained from the Geobase database using Canadian Digital Elevation Data at a scale of 1:50,000. MESH requires seven meteorological forcing variables; precipitation, atmospheric specific humidity, atmospheric temperature, atmospheric pressure, incoming shortwave radiation, incoming longwave radiation, and atmospheric wind speed. In this study, all forcing data except precipitation were extracted from the Global Environmental Model (GEM) numerical weather prediction model (*Côté et al.,* 1998). GEM has a spatial discretization of 15km and a time interval of 1 hour. Precipitation data were extracted from the Canadian Precipitation Analysis (CaPA) (*Mahfouf et al.,* 2007), a set of spatially interpolated data combining GEM precipitation and observed data from precipitation gauges. CaPA has a resolution of 15 km and produces 6 hourly rainfall data.

The soil texture data, obtained from the Soil Landscape of Canada, were used to define the feasible ranges of soil parameters. **Fig. 2.1c** shows the soil parent material texture groups in the two basins. The region is mainly dominated by moderately fine textured soils such as silty clay loams and clay loams, followed by medium coarse soils such as sandy loam soil and very coarse texture soils such as sand and loamy sand. Land cover data, obtained from the Canada Center for Remote Sensing, were used to define Group Response Units (GRUs). The land cover map of the Battle and Vermilion Basins is shown in **Fig. 2.1d**. The majority of the basin is covered by cropland, followed by grassland, and forestland.


Canada
Saskatchewan Basin
Study Area
Latitude
Longitude
80
70
60
50
40
-140
-130
-120
-110
-100
-90
-80
-70
-60
-50


b)
Basin Boundary
River
Gauging Station
Vermilion
Battle Basin
05EE007
05FE004
Latitude
Longitude
Elevation (m)
54
53.5
53
52.5
52
-114
-113
-112
-111
-110
-109
-108
1100
1000
900
800
700
600
500


c)
Coarse
Moderately fine
Moderately coarse
Fine
Medium
Very coarse
Undifferentiated
Latitude
Longitude
54
53.5
53
52.5
52
-114
-113
-112
-111
-110
-109
-108

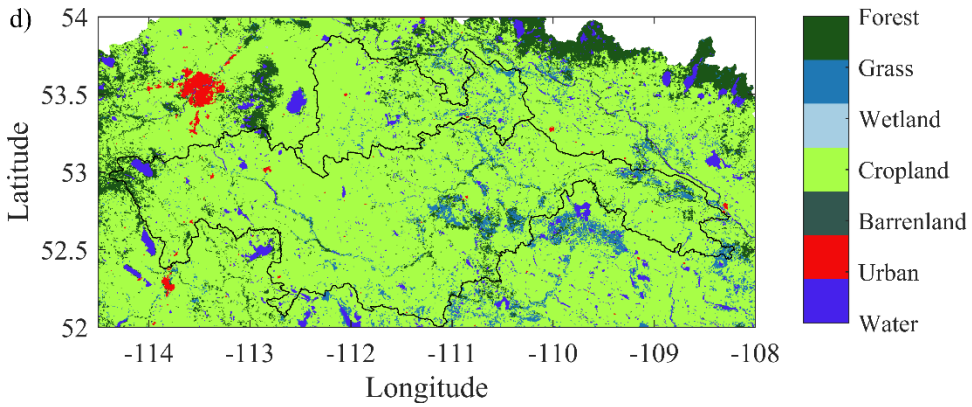


Figure 2. 1 (a) Map of Canada showing the location of the Saskatchewan River Basin and the study areas, (b) Battle and Vermilion subbasins DEM, boundary, and gauging stations, (c) the map of soil parent material texture groups, and (d) the map of landcover

2.4.2 Model implementation

MESH is Environment and Climate Change Canada's Land Surface-Hydrology Modelling System (*Pietroniro et al.*, 2007) and has been widely used in different parts of Canada (*Davison et al.*, 2006; *Mekonnen et al.*, 2014; *Pietroniro et al.*, 2007; *Haghnegahdar et al.*, 2014; *Pohl and Marsh*, 2006; *Mengistu and Spence*, 2016). MESH is a grid-based modelling system composed of three components: (1) the Canadian Land Surface Scheme (CLASS) (*Verseghy*, 1991; *Verseghy et al.*, 1993) that computes the energy and water balances using physically-based equations for soil, snow, and vegetation canopy at a half hourly time step, (2) lateral movement of soil and surface water to the drainage system with either of the algorithms called WATROF (*Soulis et al.*, 2000 or PDMROF (*Mekonnen et al.*, 2014), and (3) hydrological routing using WATFLOOD (*Kouwen et al.*, 1993) that collects overland flow and interflow from each grid cell at each time step and routes them through the drainage system. The river routing is based on a storage routing technique in which the channel roughness and storage characteristics control the inflows from the local grid and upstream river reach.

The soil was conceptualized to be represented by three layers with depths of 0.1, 0.25, and 3.75 m for the top, middle, and bottom layers, respectively. To simulate lateral flows, we used the PDMROF algorithm (*Mekonnen et al.*, 2014) which has been designed to simulate the complex

hydrologic behaviour of prairie regions. PDMROF uses the probability-density-model concept of *Moore* (2007) to parsimoniously represent the dynamic-contributing-area nature of prairies. PDMROF has two parameters. The first parameter represents the maximum storage capacity in the tile (CMAX), while the second parameter is a shape factor (B) that controls the degree of spatial variability or distribution of storage capacity (*Mekonnen et al.*, 2014; *Mengistu and Spence*, 2016). In this study, the MESH model was setup at a grid resolution of 0.125°, resulting in 220 and 68 grid cells for the Battle and Vermilion Basins, respectively. The terrain analysis was conducted at a finer resolution of 1:50,000, and the heterogeneity upscaled to 0.125° grids using the tools provided in the GIS-based GreenKenue software for watershed delineation and parametrization (*Canadian Hydraulics Centre*, 2010; *Mengistu et al.*, 2016). Heterogeneity in a grid with respect to land cover (**Fig. 2.1d**) was represented through the Grouped Response Units (GRU) concept (*Kouwen et al.*, 1993). Energy and water balances are computed at the GRU level and then aggregated to the grid scale, by weighted averaging based on GRU-to-grid-area fractions. The time resolution of all the forcing variables was transformed to half-hourly by linear interpolation to be fed to the MESH model. The effective land surface parameters (e.g. saturated hydraulic conductivity) are assumed to be the same for each land-cover type. Three families of parameters were calibrated: vegetation parameters, soil parameters, and drainage and routing parameters. **Table 2.1** lists 50 MESH model parameters and their ranges. These parameters were used in the GSA to directly characterize the significance of each individual parameter. For calibration, however, we tied the soil parameters together through Eq.s 4-7 developed by *Cosby et al.* (1984), where the soil parameters are estimated by percentages of sand and clay in the soil. Then, we used these percentages directly as calibration parameters, thereby reducing the number of parameters for calibration to 32 (14 parameters plus 18 new parameters also listed in **Table 2.1**). Not only does this "tying" technique reduce the number of calibration parameters, but also it simplifies the calibration problem, as the physical correlation between some of the actual soil parameters is directly preserved within these equations and does not have to be dealt with in calibration. Eq.s 8 and 9 were used to estimate hydraulic conductivity and moisture characteristics at different soil

depths. Interested readers are referred to *Cosby et al.* (1984), and *Verseghy* (1991) for details of the fundamentals and assumptions in these equations and how the soil moisture calculations impact the thermal calculation in the model.

$$\theta_p = (-0.126X_{sand} + 48.9)/100 \quad (4)$$

$$k_{sat} = 7.0556 \cdot 10^{-6} \exp(0.0352X_{sand} - 2.035) \quad (5)$$

$$b = 0.159X_{clay} + 2.91 \quad (6)$$

$$\psi_{sat} = 0.01 \exp(-0.0302X_{sand} + 4.33) \quad (7)$$

$$k(z) = k_{sat}\left[\frac{\theta_l(z)}{\theta_p}\right]^{(2b+3)} \quad (8)$$

$$\psi(z) = \psi_{sat}\left[\frac{\theta_l(z)}{\theta_p}\right]^{(-b)} \quad (9)$$

where X_{sand} and X_{clay} are percentages of sand and clay, respectively, θ_p is pore volume fraction, k_{sat} is saturated hydraulic conductivity, b is the slope of the retention curve, ψ_{sat} is effective "saturated" soil water suction, $k(z)$ is hydraulic conductivity, and $\psi(z)$ is soil water suction at depth z. The feasible parameter ranges shown in **Table 2.1** were chosen using recommendations in the CLASS manual (*Verseghy*, 2009) and processing the soil map based on the U.S. Department of Agriculture (1951) textural triangles.

Table 2. 1 The MESH model parameters and their feasible ranges. Parameters 1 to 14 (upper part of the table) and 15 to 50 of the lower left part of the table are actual model parameters and were used in global sensitivity analysis. Parameters 15 to 32 of the lower right part of the table are effective parameters to tie together the actual parameters of the lower left part of the table and were used in calibration along with the first 14 parameters. Parameters considered for the RSA and VARS sensitivity analysis, and their ranges of variation

No	Parameter	Description	Range
Vegetation parameters for three land cover			
1	LAMNF	Minimum leaf area index for forest [ ]	0.5 – 1.6
2	LNZ0F	Natural logarithm of the roughness length for forest [ ]	0.0 – 0.405
3	LAMXG	Maximum leaf area index for grass [ ]	3.5 – 4.0
4	LAMNG	Minimum leaf area index grass [ ]	3.0 – 3.5
5	LNZ0G	Natural logarithm of the roughness length for grass [ ]	-3.91 – -2.5
6	LAMXC	Maximum leaf area index for cropland [ ]	4.0 – 6.0
7	LNZ0C	Natural logarithm of the roughness length for cropland [ ]	-2.52 – -2.3
Routing and PDMROF parameters			
8	WFR2	River channels roughness factor	0.3 – 1.0
9	BF	Shape factor parameter for the Pareto distribution function (PDMROF) forest [ ]	0.2 – 1.5
10	CMAXF	Maximum storage parameter [m] for the pareto distribution function (PDMROF) forest [m]	1.01 – 2.0
11	BG	Shape factor parameter for the Pareto distribution function (PDMROF) grass [ ]	0.2 – 1.5
12	CMAXG	Maximum storage parameter [m] for the pareto distribution function (PDMROF) grass [m]	1.01 – 2.0
13	BC	Shape factor parameter for the Pareto distribution function (PDMROF) cropland [ ]	0.2 – 1.5
14	CMAXC	Maximum storage parameter [m] for the pareto distribution function (PDMROF) cropland [m]	1.01 – 2.0

Soil parameters for three layers and three landcover

Soil Parameters				Calibration Parameters			
No	Parameter	Description	Range	No	Parameter	Description	Range
Soil hydraulic parameters over Forestland for three soil layers							
15 – 17	θ_{pF}	Pore volume fraction [$m^3\ m^{-3}$]	0.413 – 0.464	15 – 17	X_{sandF}	% of sand [%]	20 – 60
18 – 20	b_F	Slope of retention curve [ ]	6.88 – 9.27				
21 – 23	ψ_{sF}	Effective saturated soil water suction [m]	0.124 – 0.415				
24 – 26	K_{sF}	Saturated hydraulic conductivity [$m\ s^{-1}$]	$1.864*10^{-6}$ – $7.62*10^{-6}$	18 – 20	X_{clayF}	% of clay [%]	25 – 40
Soil hydraulic parameters over grassland for three soil layers							
27 – 29	θ_{pG}	Pore volume fraction [$m^3\ m^{-3}$]	0.407 – 0.463	21 – 23	X_{sandF}	% of sand [%]	20 – 65
30 – 32	b_G	Slope of retention curve [ ]	5.29 – 7.68				
33 – 35	ψ_{sG}	Effective saturated soil water suction [m]	0.106 – 0.415				
36 – 38	K_{sG}	Saturated hydraulic conductivity [$m\ s^{-1}$]	$1.864*10^{-6}$ – $9.086*10^{-6}$	24 – 26	X_{clayG}	% of clay [%]	15 – 30
Soil hydraulic parameters over cropland for three soil layers							
39 – 41	θ_{pC}	Pore volume fraction [$m^3\ m^{-3}$]	0.432 – 0.483	27 – 29	X_{sandC}	% of sand [%]	5 – 45
42 – 44	b_C	Slope of retention curve [ ]	7.20 – 9.27				
45 – 47	ψ_{sC}	Effective saturated soil water suction [m]	0.195 – 0.653				
48 – 50	K_{sC}	Saturated hydraulic conductivity [$m\ s^{-1}$]	$1.099*10^{-6}$ – $4.494*10^{-6}$	30 – 32	X_{clayC}	% of clay [%]	27 – 40

2.4.3 Water storage data

The GRACE satellites were launched in 2002 to provide data describing the temporal change of the Earth's gravity field (geoid) approximately at a monthly time interval (*Swenson*, 2002; *Tapley et al.,* 2004). The changes in gravity field observations have been processed to estimate changes in integrated water storage at each grid cell (1°∗1°) over time. We used GRACE TWS anomaly data processed by Natural Resources Canada using a two–step filtering approach (*Huang et al.*, 2012; *Lambert et al.,* 2013) at 1°∗1° resolution for the period of 2003-2011 (except June 2006, January 2011, and June 2011 that are missing), where the mean of the GRACE product between January 2003 to December 2009 was used to estimate TWS anomaly. The standard error in the GRACE-TWS data is estimated to be about 25 mm at the basin scale (*Huang et al.,* 2012; *Lambert et al.,* 2013).

In this study, the differences between the spatial and temporal resolutions of the model and the GRACE data were reconciled by calculating and comparing their monthly basin-scale average total water storage anomalies. To this end, we looked at the basin scale monthly TWS anomaly for calculating NSE (TWS). The simulated TWS anomaly was obtained by calculating the monthly simulated TWS minus the average of TWS time series for 2003-2009 (consistent with GRACE-TWS product).

Here we assumed that the variability in GRACE-TWS is primarily attributable to the variability in surface and near-surface subsurface storages excluding possible variabilities in groundwater. This is a valid assumption in the study area, where aquifers are limited, groundwater recharge rates are low and groundwater withdrawal is overall insignificant.

2.5 Results and discussion

2.5.1 Parameter sensitivity analysis

Global sensitivity analysis was conducted on the MESH model soil, routing, vegetation, and ponding parameters. **Fig. 2.2a** and **Fig. 2.2b** show the results of the VARS-based assessment of

sensitivity of the MESH model parameters based on the Nash-Sutcliffe Efficiency on daily streamflows (NSE(FL)) and monthly GRACE-based TWS (NSE(TWS)). VARS was enabled with bootstrapping to ensure the stability of the algorithm and to assess the confidence in the results. The relatively narrow confidence intervals on this plot (generated by bootstrap) is an indication that the GSA algorithm has stabilized with the taken sample (model runs). It also indicates a high reliability of the GSA results.

Fig. 2.3 is an alternative presentation of the sensitivity metrics shown in **Fig. 2.2** to better understand the multi-criteria nature of sensitivity. This figure is a scatter plot of the sensitivity of NSE(FL) to the model parameters versus the sensitivity of NSE(TWS) to the same parameters. Evidently, the river roughness coefficient (WFR2), which is a parameter that combines channel shape, width to depth ratio and Manning's roughness, is one of the most dominant controls for NSE(FL), whereas, it is completely insensitive for NSE(TWS). This is because WFR2 controls the timing of river flows (highly influential on NSE(FL)) and has minimal impact on storage across the basin. This observation is consistent with that of *Razavi and Gupta* (2016b) on MESH and the work of *Qiao et al.,* (2013) in which the Manning's coefficient of the SWAT model was shown to be the most influential factor in controlling the quality of NSE (FL). The PDMROF parameters of cropland, BC and CMAXC, are also dominant controls on NSE(FL), while their effect is less significant for NSE(TWS). The soil hydraulic parameters of croplands are the most dominant controls of NSE(TWS) and are also highly influential on NSE(FL). In the case of NSE(TWS), the majority of the soil hydraulic parameters found to be influential in controlling the total water storage dynamics, particularly, the third soil layer's hydraulic parameters are more influential than those of the first and second layers because the third layer has a larger storage capacity.

The vegetation parameter LAMNG (minimum leaf area index for grass) has almost no influence on the model performance with respect to either of the two performance metrics, while LAMXC (maximum leaf area index for croplands) is identified to be among the influential parameters with respect to NSE(TWS). This is consistent with the fact that the majority of the basin is cropland, and also suggests that maximum leaf area index can have significant control on

the water storage in the basin. The soil parameters of grassland and forestland have very similar influence with respect to NSE(TWS). In regard to NSE(FL), however, the soil parameters of forestland tend to be more influential compared with grassland soil parameters. This indicates that forestland has a more major role in controlling the streamflow (in particular in terms of timing and peaks). However, the soil parameters of both forestland and grassland are dominated by cropland soil parameters, because cropland has a large area coverage.

The analysis above showed that the streamflow-based objective function has limited identifiability power for a sub-set of parameters (i.e., the insensitive parameters with respect to this objective function). The TWS-based objective function, however, could provide better identifiability power for some of those, as they became sensitive with respect to the new objective function. This demonstrates that streamflow and TWS data are complementary for model parameter identification, and their conjunctive use in a multiple criteria framework can potentially improve parameter identifiability.

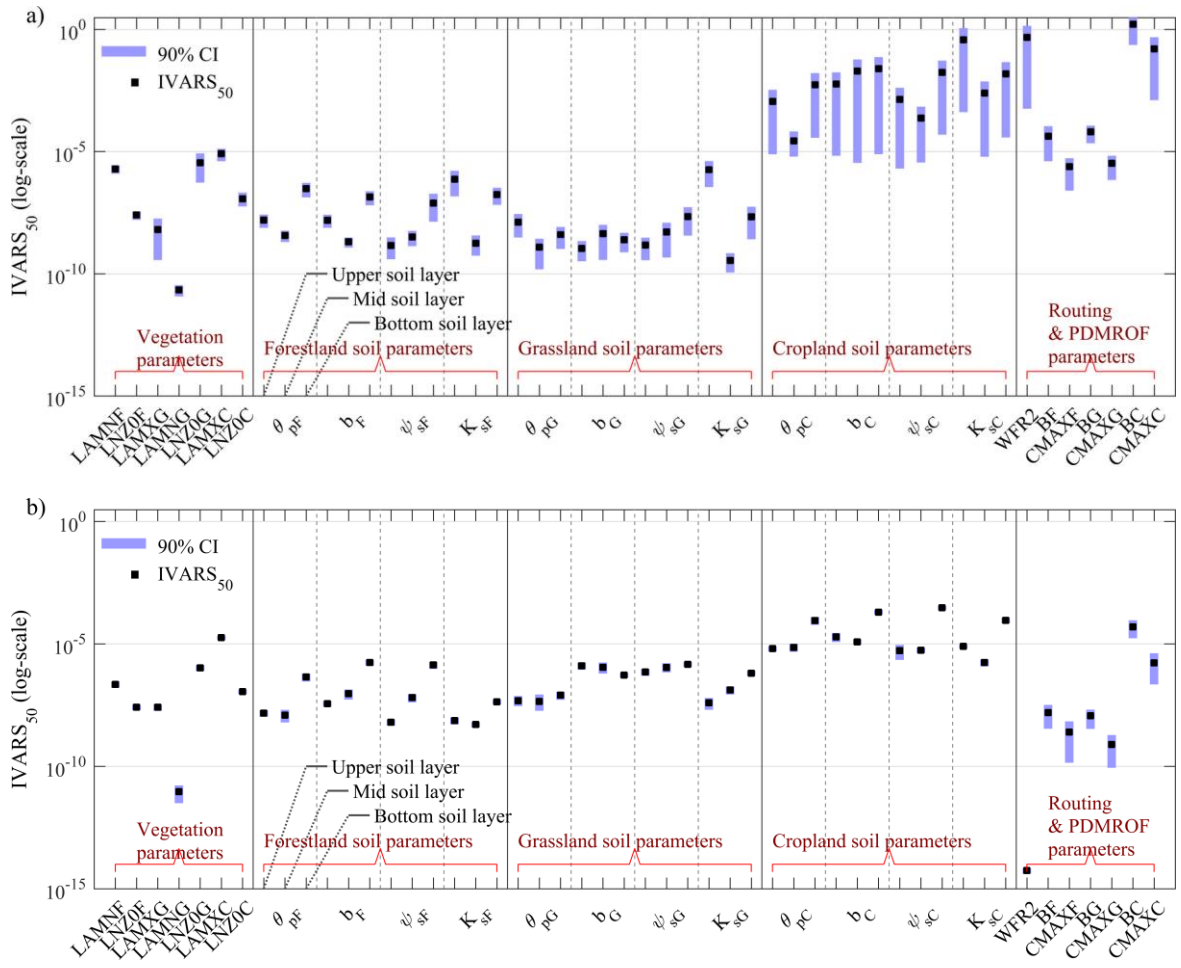


Figure 2. 2 Battle Basin results of global sensitivity analysis of model parameters using the VARS framework based on (a) NSE(FL) (flow) and (b) NSE(TWS) (total water storage) – a larger value of $IVARS_{50}$ indicates a higher rate of global sensitivity of the objective function to the associated parameter.

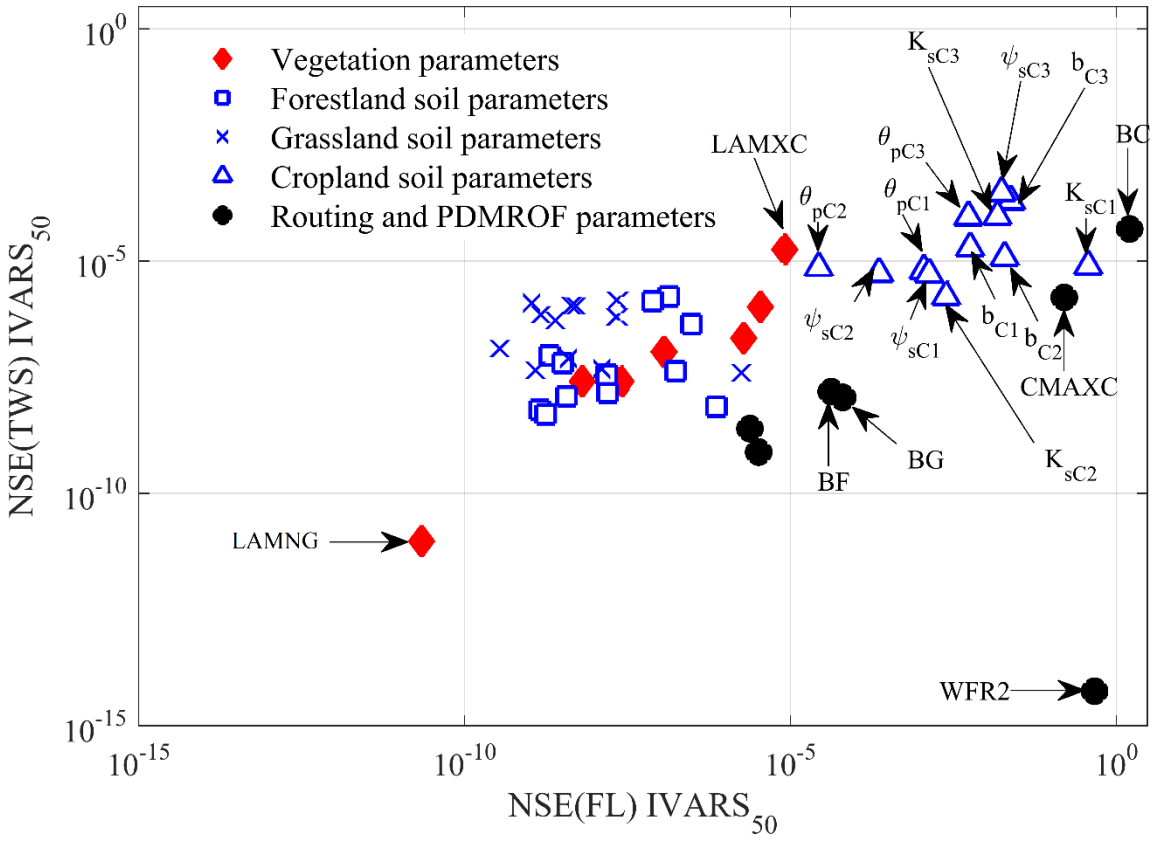


Figure 2. 3 The scatter plot of global sensitivity of NSE (TWS) to model parameters versus the global sensitivity of NSE (FL) to the same parameters.

2.5.2 Multi-criteria calibration

Fig. 2.4 shows the results of approach 1 where three streamflow-based objective functions, NSE(FL), NSE(logFL), and PBIAS(FL), were used. The best solution found in terms of NSE (FL) had NSE(FL) = 0.798, while NSE(logFL) = 0.645 and PBIAS(FL) = 3.110. However, the best NSE(logFL) was 0.790 coinciding with NSE (FL) = 0.728 and PBIAS (FL) 4.730. The best PBIAS(FL) was found to be 1.503 for a solution with NSE(FL) = 0.787 and NSE(logFL) = 0.736.

Evidently, there are significant trade-offs between the three objective functions. The competition between NSE(FL) and NSE(logFL) formed a smooth convex Pareto front (**Fig. 2.4b**).

Fig. 2.5 shows the results of approach 2, which is a four-objective optimization problem with the new objective function, NSE (TWS), and the above three streamflow-based objective functions. **Fig. 2.5a** shows the 4-dimensional Pareto front directly resulted from the experiments, and **Fig. 2.5b** and **Fig. 2.5c** look at the same Pareto front from two different angles (2-dimensional projections of the original Pareto front). To provide a comprehensive view of the entire underlying Pareto front, **Fig. 2.5d** combines the parameter sets obtained from approach 1 and approach 2. In theory, the true Pareto front of approach 1 is a subset of the true Pareto front of approach 2, indicating that those parameter sets of approach 1 that remain non-dominated in the multi-objective optimization problem of approach 2 are also essentially part of the Pareto front of Approach 2. In other words, both approaches should reach practically similar Pareto solutions in terms of the objective functions defined on streamflow (i.e., NSE (FL), NSE (logFL) and PBIAS (FL)). For example, if we map the 4-D Pareto solutions of **Fig. 2.5a** onto its NSE (FL) and NSE (logFL) 2-D sub-space (not shown), we will see the Pareto front attained by approach 1 as mapped in **Fig. 2.4b**. This means Approach 2 contains the solutions of approach 1, while adding one more dimension where NSE (TWS) is maximized as well. This new dimension in approach 2 pushes for solutions that fit TWS better, while in approach 1, the model performance with respect to TWS is not accounted for at all; That is why on **Fig. 2.5d**, approach 1 shows an inferior performance with respect to TWS. Obviously, NSE (TWS) introduced an important new angle to the evaluation of model performance, demonstrating a significant trade-off between different presumably acceptable solutions based on the streamflow-based criteria. According to our results, the NSE (TWS) could be as low as 0.45 (or lower) for the parameter sets that are deemed acceptable based on the streamflow-based objective functions (for example, assuming the thresholds of acceptability are NSE (FL)>0.75, NSE(logFL)>0.75, and PBIAS(FL)<5%). The inclusion of NSE (TWS) in optimization, however, could significantly improve the quality of solutions by increasing NSE (TWS) to be greater than 0.75 while preserving the quality in terms of the other objective functions.

For example, **Fig. 2.5d** reveals that for any NSE (FL) value greater than 0.75, the possible range of NSE (TWS) can be wide, from poor (~0.45) to high quality (~0.84). Further, the results showed that there is a significant trade-off between PBIAS (FL) and NSE (TWS), indicating that PBIAS (FL) is not sufficient to represent the water storage in the basin. Approach 2 identified parameter sets that demonstrated improved model performance with respect to all the four objective functions (error metrics).

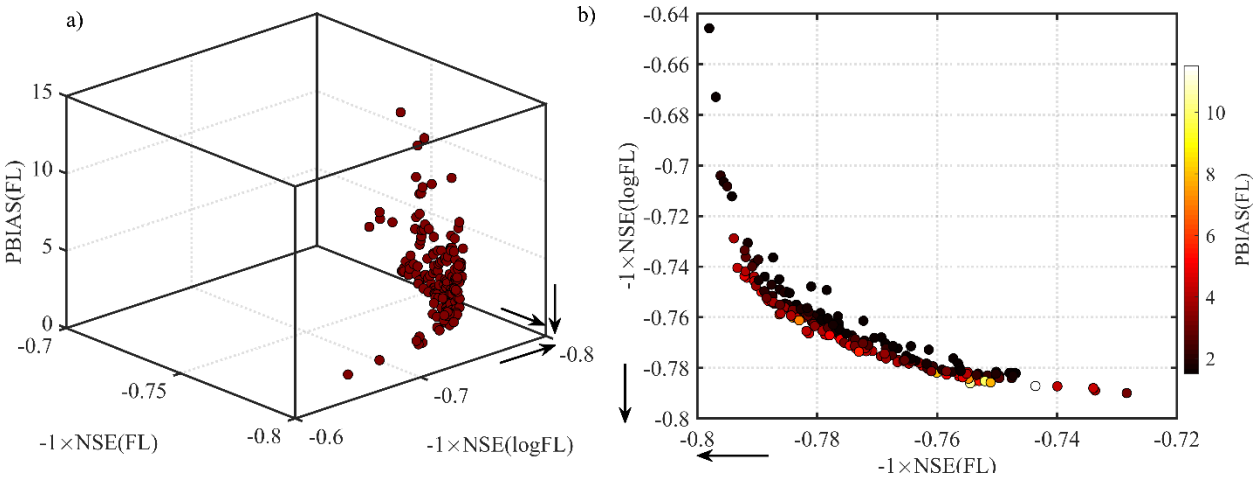


Figure 2. 4 The result of model calibration via approach 1 where three objective functions on streamflow were optimized. (a) The full 3-dimensional Pareto front between the three objective functions. (b) A 2-dimensional projection of the Pareto front onto NSE (FL) vs. NSE (logFL) space while color represents PBIAS (FL).

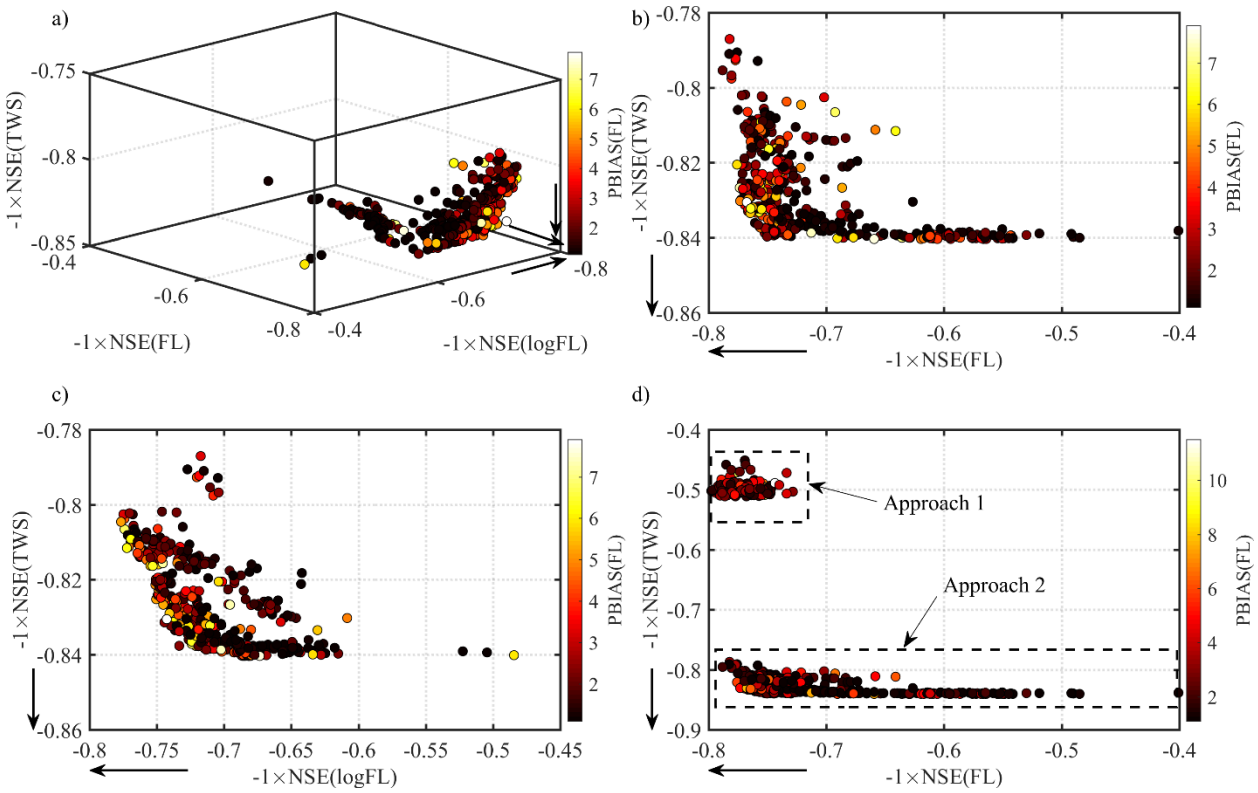


Figure 2. 5 The result of model calibration via approach 2 where four objective functions including the new objective function on total water storage were optimized. (a) The full 4-dimensional Pareto front where the fourth dimension is represented by color. (b) and (c) Selected 2-dimensional projections of the full Pareto front to better show the trade-off between the new objective function with the conventional streamflow-based objective function. (d) Comparison of the clouds of points of the Pareto parameter sets obtained by approach 1 and approach 2 in a 2-dimensional space.

2.5.3 Parameter identifiability

To assess how parameter identifiability improves as a result of calibration to GRACE-TWS in addition to streamflows, we screened and compared the behavioral parameter sets obtained via the two approaches by defining an arbitrary screening threshold on model evaluation metrics to select well-behaving parameter sets and simulations. Hence, the behavioral parameter sets were defined to be the parameter sets lying on the Pareto fronts of each replicate of each optimization experiment which satisfied the following constraints: PBIAS (FL) $\leq$ 10% and NSE (FL) $\geq$0.6 and NSE (logFL) $\geq$0.6 for approach 1 and the aforementioned criteria plus NSE (TWS) $\geq$0.6 for approach 2. For this analysis, the calibration parameters for percentages of sand and clay in soil in the three soil layers were transferred back to actual model parameter values (see Section 2.4.2 for details). **Fig. 2.6** shows the range of behavioral values of cropland soil hydraulic parameters as well as routing parameters, scaled between zero and one. These parameters were identified to be influential with respect to streamflow and/or TWS in section 2.5.1.

In general, the behavioural parameter space in approach 2 was more constrained (smaller) compared with the behavioural parameter space in approach 1; the parameter ranges of approach 2 were narrower than (or in some cases similar to) those of approach 1. This indicates that approach 2 has a higher identifiability power for most of the parameters. This is in particular the case for the soil parameters in the second and third soil layers which were among high sensitive parameters to GRACE storage. The higher identifiability power of approach 2 allowed us to exclude solutions that seem to be acceptable in approach 1. In terms of the routing parameters (WFR2, BC, CMAXC), however, approach 2 could not improve the identifiability much (not much reduction in their identified ranges). Consistent with the results of the sensitivity analysis, this behaviour was expected, as routing parameters have minimal impact on TWS in comparison with the soil parameters.

The identified behavioral parameter ranges for approach 1 and 2 showed a similar region of identification for most of the parameters, as shown in **Fig. 2.6** (normalized parameter space). However, notable differences observed in the identification of behavioural regions for the

parameters of the third soil layer (θ_{p3}, ψ_{sat3}, k_{sat3}, and b_3). These third soil layer parameters (see Section 2.5.1 for their description) control the amount of water stored and how fast stored water drains from the third soil layer. Since the third soil layer has a larger depth, its parameterization significantly impacts the water storage dynamics over time, while its impact on streamflow is less as the streamflow metrics showed relatively more sensitivity to the top soil layer parameters. As presented in the next section (Section 2.5.4), the difference in the regions of identified parameters can introduce different trajectories and patterns of how the water storage evolves over time. Compared with previous studies that showed the benefit of GRACE via an aggregated multi-objective approach (e.g., *Qiao et al.*, 2013; *Lo et al.*, 2010) or a fully multi-objective approach (e.g., *Werth and Güntner*, 2010; *Livneh and Lettenmaier*, 2012), our results demonstrate the benefit of the proposed approach in parameter identification.

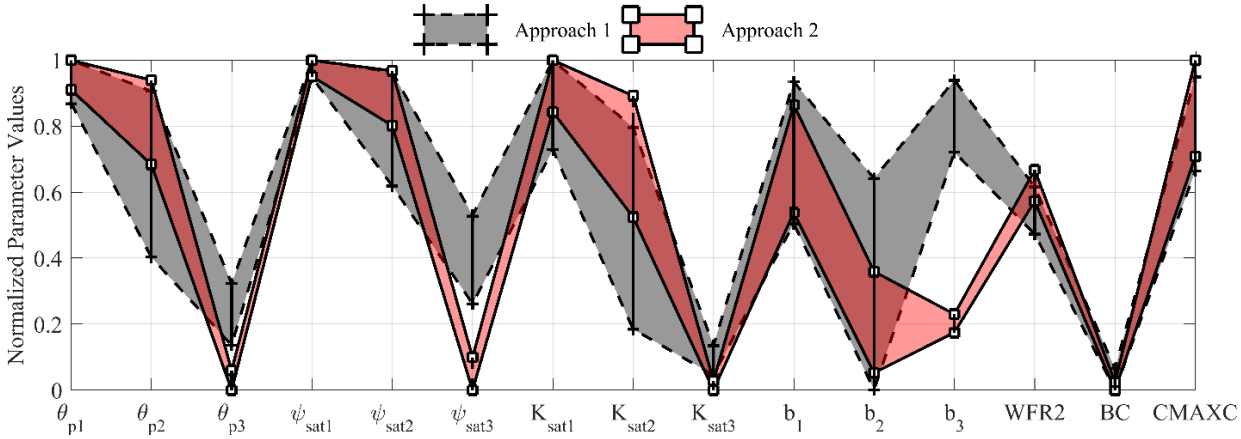


Figure 2. 6 A representation of parameter identifiability via approach 1 and approach 2. Only the most sensitive parameters selected based on global sensitivity analysis are shown.

2.5.4 Simulations in calibration and validation

Fig. 2.7a and **Fig. 2.7b** show the envelopes of simulated hydrographs for the Battle River in calibration and validation (in time) obtained via the behavioural parameter sets (identified in Section 2.5.3) of the two approaches along with observations. The envelopes of both approaches

are comparable and relatively narrow. For the envelope shown of approach 1, the best NSE (FL), NSE (logFL), and PBIAS (FL) are 0.80, 0.79, and 1.50, respectively during calibration, and 0.68, 0.78, and 11.5, respectively during validation. The same figures of approach 2, however, are 0.79, 0.78, and 1.1 during calibration, and 0.70, 0.68, and 3.5 respectively during validation. Overall, the two approaches performed comparably in reproducing streamflows in calibration. In validation, the behavioral solutions obtained by approach 2 better reproduced high-flows and the total volume of water, while the ones by approach 1 leaned towards better representing low flows (unlike calibration). We note that the inferior performance of approach 2 in terms of reproducing low flows in validation occurred mainly in 2010. Although the obtained (optimized) error metrics are in the typically acceptable range in the field, the quality of simulated hydrographs may be viewed as sub-standard for some time periods. Notably, however, this is arguably the current best modelling practice in prairie hydrology, given the hydrologic complexity of prairie regions manifested in their fill-and-spill and variable-contributing-area mechanisms of the model (for detail see *Mekonnen et al.* 2014).

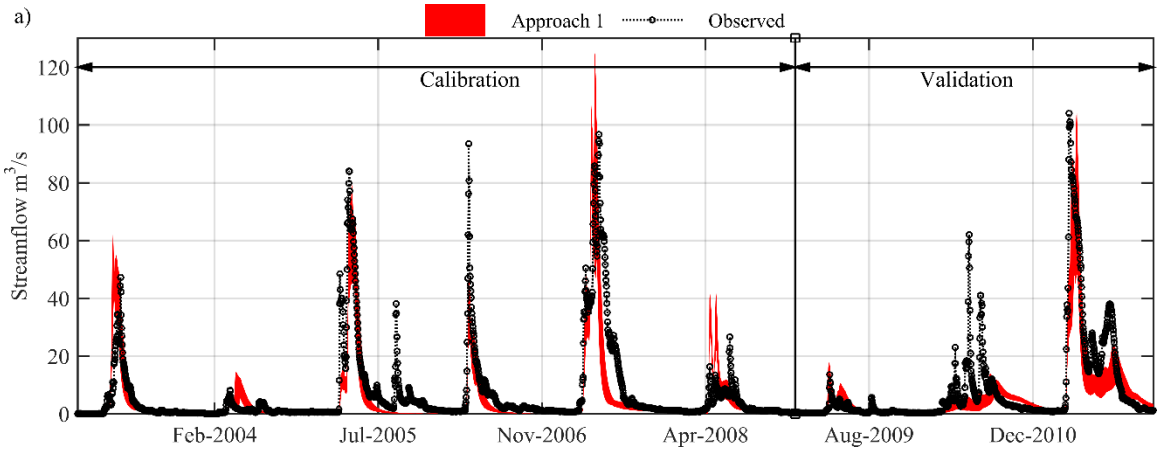
We also utilized two streamflow-based signatures for better characterization of simulated streamflows, the Slope of Flow Duration Curve (SFDC) and Peak Distribution (PD) [for their formulation see *Sawicz et al.* (2011) and *Euser et al.* (2013), respectively]. The SFDC for the observed streamflow in the calibration and validation periods is 3.71 and 4.48, respectively. The ranges of SFDC in the envelope of streamflows obtained by approach 1 in the calibration and validation periods are [5.33-6.15] and [4.35-5.47]. These ranges for the envelope associated with approach 2 in the calibration and validation periods are [2.70-3.38] and [3.45-4.18]. The envelope of approach 1 largely over-estimated the observed SFDC in calibration and to a lesser extent in validation. The envelope of approach 2 slightly under-estimated the observed SFDC in both calibration and validation. On the other hand, the PD for the observed streamflow in the calibration and validation periods is 45.3 and 57.6, respectively. The ranges of PD by approach 1 in the calibration and validation periods are [20.4-34.2] and [15.9-29.9]. These ranges for approach 2 in the calibration and validation periods are [27.9-39.3] and [34.4-44.0]. As such, the envelope of

approach 1 largely under-estimated the observed PD in both calibration and validation, while the ones of approach 2 underestimated the observed PD but to a considerably lesser extent in both calibration and validation. The streamflow simulation results for the Vermilion River for period 2003-2011 (validation in space) (**Fig. 2.8a** and **Fig. 2.8b**) were quite consistent with the results presented above for the Battle River.

Overall, based on the results across the two watersheds, it could be concluded that approach 1 and approach 2 were comparably capable of reproducing streamflows, with approach 2 slightly outperforming approach 1 (given all criteria discussed). However, in terms of TWS, this assessment is significantly different. **Fig. 2.9a** and **Fig. 2.9b** show the envelopes of simulated TWS of the same behavioral parameter sets for the two approaches in the Battle and Vermilion Basins. Unlike the result of approach 1 in the Battle Basin (calibration and validation in time), the envelope generated by approach 2 consistently followed the GRACE-based TWS. The width of this envelope is significantly narrower than that of approach 1 across the time domain. Likewise, the simulated TWS by approach 2 in the Vermilion basin (validation in space and time) is in agreement with GRACE-TWS, whereas they deviate significantly from each other in approach 1. These demonstrate that approach 2 is superior in reproducing TWS in both calibration and validation.

Notably, the performance of the model calibrated via approach 1 may be deemed very poor (although promising in terms of streamflows), as this model seems to falsely demonstrate a significant increasing trend in TWS across the basin. This indicates the inadequacy of the information content of streamflows to guide towards understanding other fluxes and to effectively constrain the model's degree of freedom. Importantly, this type of behavior can occur in any land surface and hydrology models that have a thick bottom soil layer in which the soil parameterization can significantly affect the dynamics of soil moisture storage and its trajectory through time. Our results highlighted the need and importance of calculating and analyzing the change in total water storage for enhanced model identification. This has been rarely reported in the literature for modelling studies, while it is particularly essential when evaluating the added value of new information in hydrologic modelling.

Finally, **Fig. 2.10** looks more closely at the behavioural parameter sets obtained in calibration and their performance in validation in space in the Vermilion basin, where the scatter plot of NSE (FL) versus PBIAS (FL) for both approaches is shown. The value of NSE (FL) varies in the range 0.41 to 0.61 in approach 1, whereas the same figure varies from 0.49 to 0.59 for approach 2. The superiority of approach 2 over approach 1 is more significant in terms of PBIAS (FL), as it varies between 15 to 45% in approach 1, but between 12 to 34% in approach 2. Most importantly, in terms of TWS (shown in **Fig. 2.10** by color), approach 2 could achieve NSE (TWS) values ranging between 0.44 to 0.60 (for different behavioural parameter sets obtained in calibration), whereas approach 1 even failed to produce a positive NSE(TWS) value (all ranging between -0.43 and 0).



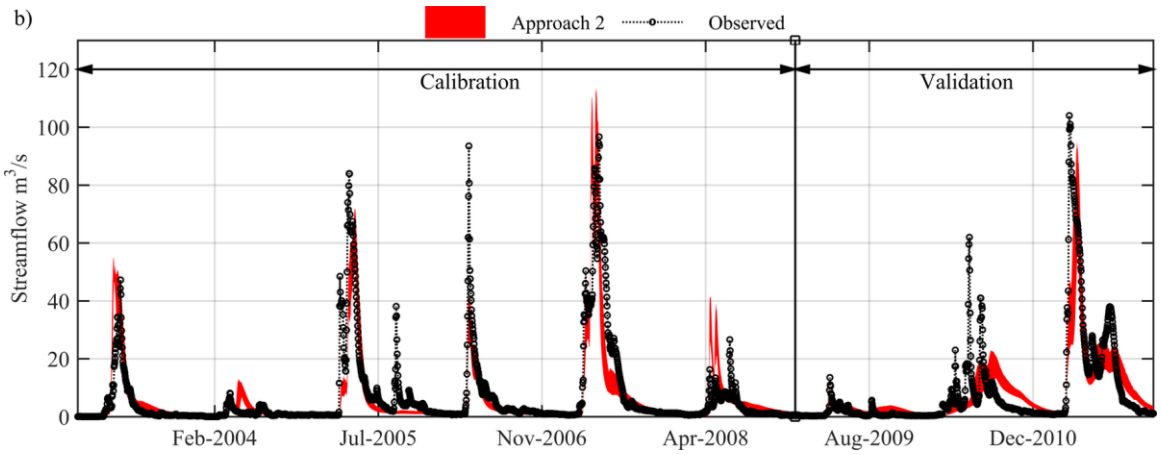


Figure 2. 7 The envelope of daily simulated streamflows of the Battle River basin obtained from all behavioural parameter sets on the Pareto fronts of (a) approach 1 and (b) approach 2, along with the daily observed streamflows.

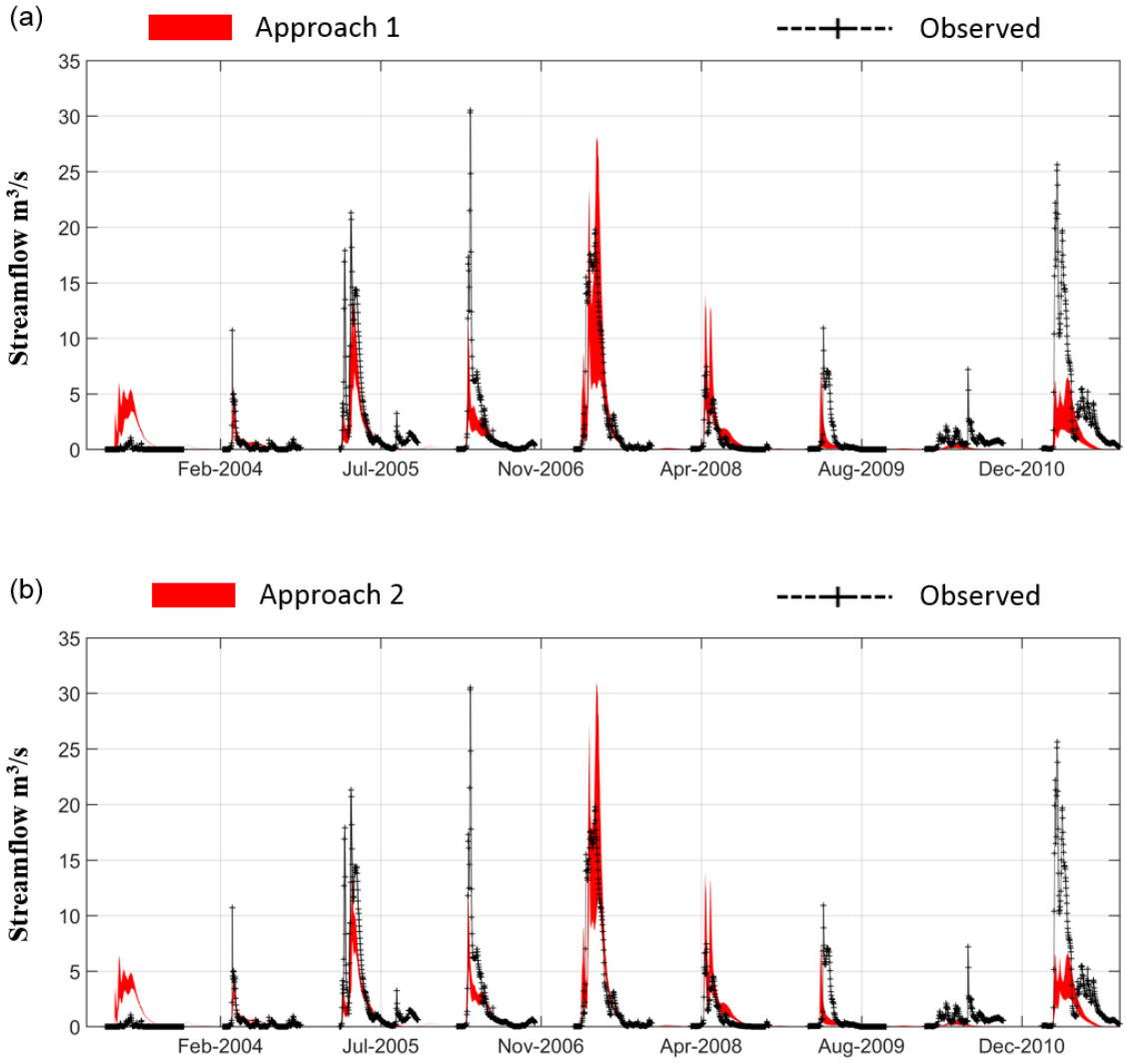


Figure 2. 8 The envelope of daily simulated streamflows of the Vermilion River basin obtained from all behavioural parameter sets on the Pareto fronts of (a) approach 1 and (b) approach 2, along with the daily observed streamflows.

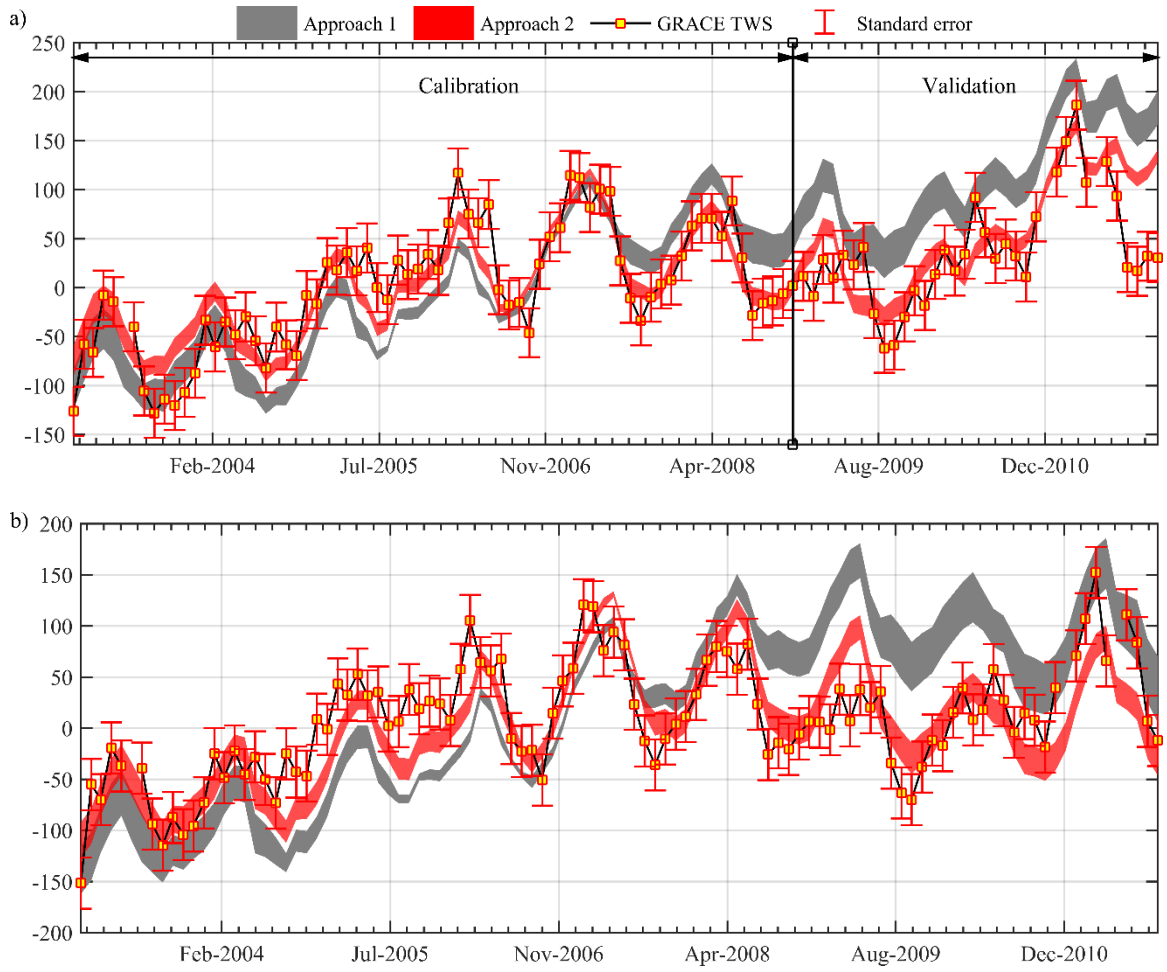


Figure 2. 9 The envelopes of monthly simulated total water storage (TWS) anomaly (mm) of the (a) Battle River and (b) Vermilion River Basins obtained from all behavioural parameter sets on the Pareto fronts of approach 1 and approach 2, along with GRACE-TWS anomaly (mm) - the standard error of GRACE-TWS is plotted for demonstration only and not used in any calculation.

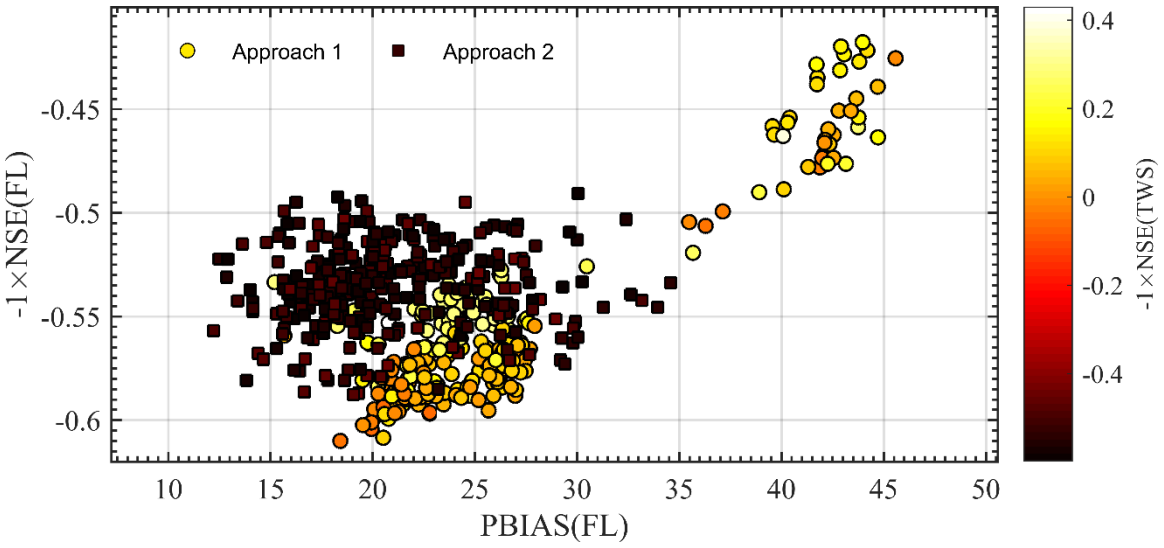


Figure 2. 10 The scatter plot of Nash-Sutcliffe Efficiency on streamflows NSE (FL) versus Percent Bias PBIAS (FL) in validation (Vermilion River basin) for all the behavioral parameter sets obtained via approach 1 and approach 2.

2.6 Summary and conclusions

This study argued that streamflow-oriented model development and calibration may not be capable of effectively constraining the dynamical behavior of complex models within the physically justifiable range, and reiterated the need for moving beyond calibration only to streamflows. The global sensitivity analysis of streamflow and TWS anomaly to model parameters revealed that the parameters that are dominant controls of water storage across the basin (e.g., soil properties of deeper soil layers) may be less influential on streamflows (at the typical time scale of model calibration). In contrast, the drainage and routing parameters that significantly control the streamflow and its timing were shown to be less or not influential with respect to the water storage in the basin.

The calibration experiments with and without GRACE-based data showed significant trade-offs between different errors metrics, most notably between the TWS-based error metrics

and the three streamflow-based error metrics. These trade-offs introduced a new angle to the evaluation of model performance, indicating the weakness of streamflow-based error metrics in adequately identifying model parameters. Most importantly, it was shown that minimizing volume bias of streamflows (in conjunction with other streamflow-based metrics) was not sufficient to properly capture water storage in the basin (and subsequently to close the water balance) and that the GRACE-based TWS data could address this weakness.

The trade-offs observed between the different error metrics may be partly attributable to the errors in data and model structure, as these error metrics are not inherently conflicting. That being said, a probably more important point is that these trade-offs are a characterization of the equifinality principle in dynamical systems modelling, where different system trajectories could lead to similar outcomes with respect to a given metric. In other words, the solutions on the trade-off represented a range of trajectories, each of which could be deemed acceptable with respect to one or more metrics depending on where the solution was located in the trade-off. Our multi-criteria calibration could contribute to addressing equifinality and in particular the issue of non-uniqueness of optimal parameter values with respect to streamflow-based metrics, by providing an important new dimension.

The added dimension to the problem of model calibration improved in constraining constrain the behavioral parameter space, and subsequently, the dynamical behavior of the model over time. This was evident as significant differences were observed in the time series of simulated water storage in the basin under the two calibration approaches. Importantly, with streamflow-based calibration, significant unrealistic (even false) trends were observed in the time series of simulated water storage, whereas streamflow simulations, including total streamflow volume, were adequately accurate. The addition of the GRACE-based TWS error metric was able to improve the simulation of water storage in the basin, and subsequently, the representation of the hydrologic processes.

Although our analyses were focused on two Canadian river basins, the proposed methodology and findings can be of global significance. The adequate representation of land surface processes, including soil moisture, is essential for both land surface and hydrologic applications. Given that land surface models serve as the lower boundary condition of global and regional climate models, the quality of their performance will be reflected to the simulation and prediction by climate models. This study highlighted the need to better constrain the model's degrees of freedom by looking at the model internal state trajectories. It further demonstrated the implication that more constraints to different model states and fluxes could improve parameter identifiability. Such constraints may help significantly in dealing with the equifinality issue, which is a critical feature of the current generation of hydrological models. This is particularly important for future prediction of streamflow under climate change scenarios, where the model is run to produce possible trajectories of a system's behavior for the next century. Therefore, future research may be directed at the inclusion of a range of other remotely-sensed data of state and flux variables in a unified multi-criteria global sensitivity analysis and optimization framework for watershed model development and calibration. Such an approach has the potential to be extended to the continental or global scale.

CHAPTER 3
REPRESENTATION OF WATER MANAGEMENT IN HYDROLOGICAL AND LAND SURFACE MODELS

This chapter is based on the following published article.

Yassin F, Razavi S, Elshamy M, Davison B, Sapriza-Azuri G, Wheater H. 2019. Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrology and Earth System Sciences* **23** (9): 3735–3764 DOI: 10.5194/hess-23-3735-2019 https://www.hydrol-earth-syst-sci.net/23/3735/2019/hess-23-3735-2019-metrics.html

Author contributions: FY developed the methodology together with SR and HW. FY wrote all the computer codes, coupled with MESH model, designed the experiments, and performed them all. FY was responsible for the verification of the results. FY prepared the manuscript and other co-authors contributed to editing of the paper at all stages.

Synopsis

Reservoirs significantly affect flow regimes in watershed systems by changing the magnitude and timing of streamflows. Failure to represent these effects limits the performance of hydrological and land surface models (H-LSMs) in the many highly regulated basins across the globe and limits the applicability of such models to investigate the futures states of watershed systems through scenario analysis (e.g., scenarios of climate, land use, or reservoir regulation changes). An adequate representation of reservoirs and their operation in an H-LSM is therefore essential for a realistic representation of the downstream flow regime. In this paper, we present a general parametric reservoir operation model based on piecewise linear relationships between reservoir storage, inflow, and release, to approximate actual reservoir operations. For the identification of the model parameters, we propose two strategies: (a) a "generalized" parameterization that requires a relatively limited amount of data; and (b) direct calibration via multi-objective optimization when more data on historical storage and release are available. We use data from 37 reservoir case studies located in several regions across the globe for developing and testing the model. We further build this reservoir operation model into the MESH modelling system, which is a large-scale H-LSM. Our results across the case studies show that the proposed reservoir model with both parameter identification strategies leads to improved simulation accuracy compared with the other widely used approaches for reservoir operation simulation. We further show the significance of enabling MESH with this reservoir model and discuss the interdependent effects of the simulation accuracy of natural processes and that of reservoir operations on the overall model performance. The reservoir operation model is generic and can be integrated into any H-LSM.

3.1 Introduction

Human interventions in natural hydrologic systems, through damming and storing of water, diversion, surface and groundwater abstraction, irrigation, and land use change, have significantly altered the natural river flow regimes and the terrestrial water cycle of many river basins (*Vörösmarty et al.,* 1997, 2003; *Oki and Kanae*, 2006; *Wisser et al.,* 2010; *Haddeland et al.,* 2014). These interventions are to fulfil different types of demands such as domestic, industrial, irrigation, and hydropower demands, and to meet other needs such as flood control and conservation of aquatic habitats. With a total storage volume of more than 8000 km^3 (*ICOLD*, 2003; *Hanasaki et al.,* 2006; *Vörösmarty et al.,* 2003), more than 50,000 dams have been constructed globally to regulate more than half of the world's large river systems (*Nilsson et al.,* 2005). The aggregate storage volume of these dams is greater than 20% of the global mean annual runoff (*Vörösmarty et al.,* 1997) and is three times the annual average water storage in world's river channels (Hanasaki *et al.,* 2006).

Despite the benefits in terms of enhancing water availability in support of food security, power supply, and flood control, dams result in several negative environmental and social consequences. Adverse environmental effects include changes in natural river dynamics in terms of water temperature, sediment and nutrient transport, etc. and the fragmentation and loss of biodiversity (*Vörösmarty et al.,* 2010). Reservoirs can also intensify evaporation, by increasing the surface area of water exposed to direct sunlight and air, and through water supply for irrigation (*de Rosnay et al.,* 2003; *Pokhrel et al.,* 2012). Other environmental impacts of dams include the alteration of landscapes due to dam construction and changes to land-atmosphere interaction that can have a profound impact on local/regional climate (*Hossain et al.,* 2012). Adverse social effects include the displacement of people living near the dam site, changes to fishing patterns, and downstream erosion (*Strobl and Strobl*, 2011, p. 449). There are research gaps remaining in evaluating both positive and negative social impacts of dams (*Kirchherr et al.,* 2016). Such gaps have been the subject of many studies in both academia and industry for years, and recently, have led to the formalization of the study area of "socio-hydrology" (*Sivapalan et al.,* 2012; *Sivakumar*,

2012), which addresses more generally the inter-relationships between human and natural water systems.

Dams and reservoirs change the natural flow regimes in rivers, both in terms of magnitude and timing of flows. As a result, for rivers that contain large or small dams and reservoirs, flow regimes are a combination of natural and managed flows. Various modelling communities manage this mix of natural and managed flows differently. *Archfield et al.* (2015) compare three communities of models that can be used at continental scales: catchment models (CM), global water security models (GHM), and land-surface models (LSM). CMs generally ignore water management and focus on unmanaged headwater catchments. GHMs have been utilized in global-scale streamflow simulations and generally consider large-scale water management, but are hindered by a lack of data on large-scale water management operational decisions. LSMs have traditionally focused on providing lower boundary conditions for atmospheric models, but are increasingly being used for hydrological applications in which they are referred to as Hydrologic Land Surface Models (H-LSMs). LSMs generally ignore water management (*Clark et al.,* 2015; *Davison et al.,* 2016), with a few exceptions (e.g. *Voisin et al.* 2013a, 2013b). A fourth community of water models, that is relevant to the work presented here, are water management models (WMM) (*Yates et al.,* 2005; *Labadie,* 1995). Water modellers who know how the water is managed within their basins of interest generally use WMMs (*Lund and Guzman*, 1999; *Labadie,* 2004; *Kasprzyk et al.,* 2013). These models contain very detailed representations of water management decisions, but often consider natural flow processes in a much more rudimentary fashion than CMs.

Modelling the many managed basins around the world using CMs or LSMs can result in models with limited fidelity, which raises questions concerning the credibility of their predictions of future water resources in basins with dams and reservoirs. Therefore, there is a pressing need for better characterization of the operations of dams and reservoirs and their integration into hydrological modelling frameworks using CMs and LSMs (*Nazemi and Wheater*, 2015a and 2015b; *Wada et al.,* 2017; *Pokhrel et al.,* 2016). This need motivated the objectives of this study,

described in Section 3.2, as well as previous research, outlined in Section 3.3. The integration of reservoir regulation into hydrological modelling frameworks will improve our ability to simulate highly regulated basins around the globe, leading to better understanding of historical conditions of water resource systems and improved assessment and prediction of their future vulnerability to climate and environmental change.

3.2 Objectives

Building upon previous research, this study aims to:

- Develop and test an improved reservoir operation model that can be integrated into any CM and LSM at any scale, but in particular at large scales. Of interest is a simple but effective parametrization that can adjust to varying levels of data availability.
- Integrate the developed reservoir operation model into an LSM and evaluate its performance when working in combination of other processes in the model. Also of interest is to assess the potential conceptual and technical issues in this integration.

This paper looks to improve the representation of dams and reservoirs within CMs and LSMs. Another approach would be to couple CMs and LSMs with WMMs, but that approach is not examined here due to the fact that WMMs generally require information on how water is managed within a basin, and we are particularly interested in the more generic case when this information is likely to be limited or unavailable.

The organization of the remainder of the paper is as follows. Section 3.3 reviews different existing approaches in the literature for the representation of reservoir operation in hydrologic models. Section 3.4 presents the proposed reservoir operation model and the metrics used to evaluate it, in comparison with other existing models. Section 3.5 provides a description of the reservoir dataset used for the developments and testing. Section 3.6 presents the assessment results and comparisons. Section 3.7 ends the paper with a summary of the main findings and conclusions.

3.3 Existing reservoir models in catchment models and land surface models

An adequate representation of human interventions in Earth systems models is a major challenge. Systematic approaches towards full integration are needed as outlined in the recent studies of *Nazemi and Wheater* (2015a and 2015b), *Wada et al.* (2017), and *Pokhrel et al.* (2016). In this work, our focus is on the representation of dam and reservoir operations in catchment models (CMs) and Land Surface Models (LSMs), particularly when used at large scales. While there has been tremendous progress in the last decades in modelling the operation and management of reservoir systems at local to regional scales (e.g., *Castelletti et al.,* 2010; *Chang et al.,* 2010; *Fraternali et al.,* 2012; *Razavi et al.* 2012; *Asadzadeh et al.* 2013; *Guo et al.,* 2013), a gap still exists between the methodologies applied for local/regional-scale reservoir operation and management and the representation of reservoir operations in Earth systems models, particularly in LSMs. This gap is due to a two-fold challenge. First, the upscaling of methodologies used at smaller scales to larger scales is non-trivial; and second, the availability of data on reservoir operation and water use is often limited in many parts of the world. For example, the reservoir purpose and operational details are not always known and large reservoirs typically serve several purposes (*Wisser et al.,* 2010). As a result, most current hydrological modelling activities with CMs and LSMs, if not all, offer only a limited capability in simulating reservoir operations, whereas reservoir operations in practice involve a complex set of human-driven processes and decisions.

The existing reservoir operation methods in hydrologic models can be categorized roughly into four groups, based on their level of complexity in representing flow regulation; (1) natural lake methods, (2) inflow-and-demand based methods, (3) artificial neural network techniques, and (4) target storage-and-release based methods.

3.3.1 Natural lake methods

Methods in the first category use formulations developed for the simulation of natural lakes or uncontrolled reservoirs. In these methods, the downstream release is calculated as a function of

reservoir storage characterized by some empirical parameters (*Meigh et al.,* 1999; *Döll et al.,* 2003; *Pietroniro et al.,* 2007; *Rost et al.,* 2008). For instance, *Meigh et al.* (1999) calculate the release by $Q_t = S_t^{1.5}$ where Q_t and S_t are release and reservoir storage, respectively. Their method was later modified by Döll *et al.* (2003) such that $Q_t = b_1(S_t - S_{min})(\frac{S_t - S_{min}}{S_{max} - S_{min}})^{b_2}$ where b_1 and b_2 are release coefficients, and S_{min} and S_{max} are minimum and maximum allowable reservoir storages. The advantage of this method, as shown in *Döll et al.* (2003), is its minimal data requirement, which supports its global applicability to model lakes, reservoirs and wetlands. However, it has limited functionality to adequately represent managed reservoirs due to not accounting for reservoir operation policies to constrain or increase releases at different phases of reservoir storage dynamics. Such simplistic methods ignore the fact that the operation of a reservoir depends on the reservoir purpose and the seasonal pattern of the mismatch between the demands it supports and the inflow it receives.

3.3.2 Inflow-and-demand based methods

The inflow-and-demand based methods include reservoir water balance models that determine reservoir release using a function that accounts for inflow or a combination of inflow and demands. The most basic method in this group is the method used in *Wisser et al.* (2010), which estimates the release as a function of mean annual inflow and a set of empirical parameters that can be calibrated in the absence of information on the actual operation of a reservoir.

Hanasaki et al. (2006) pioneered the development of inflow-and-demand reservoir models and laid the foundation for many subsequent developments. The method of *Hanasaki et al.* (2006) simulates reservoir release at a monthly time step within a global routing model, and accounts for water withdrawals for reservoirs categorized as irrigation reservoirs. They grouped reservoirs serving all others purposes as non-irrigation reservoirs. This approach first estimates a provisional total annual release at the beginning of the water year based on the long-term mean annual inflow adjusted by an annual release coefficient. Then, a monthly provisional release is estimated based on the purpose of the reservoir (irrigation or non-irrigation). Downstream demands are accounted

for in irrigation reservoirs only. The provisional monthly release for large reservoirs is then modified by the annual release coefficient to calculate the actual monthly release, and the provisional monthly release for small reservoirs is additionally adjusted based on the monthly inflow to calculate the actual monthly release. The release coefficient is estimated as a function of the reservoir storage at the beginning of the operational year and the reservoir capacity (the formulation of *Hanasaki et al.* (2006) is briefly explained in section 3.4). The release coefficient reduces the current year release if the storage at the beginning is low and vice-versa. Thus, the release coefficient accounts for inter-annual variability and facilitates the representation of strategies to overcome reservoir depletion in dry years and flood overtopping in wet years.

The implementation of the release coefficient is one of the limitations of *Hanasaki et al.* (2006), because it depends only on the year's initial storage and does not account for the actual inflow of the current operational year, i.e. it does not use foresight. The initial storage reflects the recent past of the operation of the reservoir, while the actual inflow could be considerably different than the long-term mean annual inflow. For instance a sequence of low flow years would result in a low initial storage while the current year inflow (which is not known yet) could be high, and vice versa. Additionally the simplification of complex reservoir operations in *Hanasaki et al.* (2006) by using the mean annual inflow and a release constraining coefficient produces errors. However, the method is generic and has low data requirements which are advantageous. The results showed that the reservoir algorithm improved monthly discharge simulation compared to the natural lake method (*Hanasaki et al.,* 2006). The approach is effective and has found wide applicability in several global hydrological and land surface models.

The original *Hanasaki et al.* (2006) reservoir model has been modified in subsequent studies to address some of it limitations. For example, it has been modified for water extraction and other reservoir functions such as fulfilling environmental flows (*Hanasaki et al.,* 2008a, b; *Pokhrel et al.,* 2012a), and been adjusted to address direct precipitation over and evaporation from the reservoir (*Döll et al.,* 2009).

Biemans et al. (2011) added new functionalities to the *Hanasaki et al.* (2006) reservoir model related to irrigation water demand and supply distribution and ran it at a daily time step. Their contributions include: 1) modifying irrigation withdrawals to account for conveyance losses and irrigation efficiency, 2) adjusting the minimum release to 10% of the mean monthly inflow, 3) prioritizing irrigation over flood control, 4) using regulated flow instead of natural flow, to estimate mean annual inflow, 5) storing the "flow to be released" for five days in the reservoir – to mimic the storage within the conveyance system – before it is released to the river. *Voisin et al.* (2013a) further modified the reservoir model of *Hanasaki et al.* (2006) to include multipurpose functionalities (irrigation and flood control) by changing the operation to release more before the onset of snowmelt-flood season so that there will be sufficient room to store flood waters form snowmelt in the reservoir. The modification requires the specification of a flood control period. *Voisin et al.* (2013a) have also evaluated the uncertainty of reservoir simulation by comparing withdrawal vs. consumptive demands, and natural vs. regulated flow for configuring operating rules. The results of *Voisin et al.* (2013a) demonstrated that adding flood control in reservoir operation, along with a parametrization using mean annual natural inflow, and mean monthly withdrawals, improves the reservoir storage and flow simulation.

Haddeland et al. (2006) developed another pioneering generic reservoir model that has been implemented in a routing model at a daily time step to study the impact of reservoir and irrigation water withdrawals on continental surface water fluxes. The model is retrospective, i.e., it assumes full knowledge of the upcoming operation-year reservoir inflow. The reservoir operation is conducted using an optimization scheme to determine the optimal release to satisfy different sectoral demands and targets that are defined in the form of objective functions. In the case of a multipurpose reservoir, the model gives priority to irrigation demand, followed by flood control and hydropower production. Minimum flow is estimated using natural flow based on seven-day consecutive low flows with a ten year recurrence period. The flood protection objective function is minimizing reservoir release above the bankfull discharge, which is estimated using the long-term mean of annual maximum discharge. Irrigation is optimized to satisfy downstream

irrigation demand, while hydropower is optimized to increase power production. Predicting inflows for the current operational year, if possible, would allow the method to optimize the release while accounting for the whole operational year, otherwise to optimize day to day release without accounting for the remaining operational year would require several constraints. The maximum daily release is set based on the reservoir water balance that sets the storage at the end of the operational year to vary between 60 to 80% of the maximum capacity.

Similar to *Hanasaki et al.* (2006), the model of *Haddeland et al.* (2006) is favourable due to its generic formulation and capability to represent multipurpose reservoirs, and to extract water for irrigation from reservoirs. These factors make the model applicable for large-scale hydrologic models, when data on operational policies are limited (*Adam et al.,* 2007; *Van Beek et al.,* 2011). One limitation of the *Haddeland et al.,* (2006) methodology is that it requires knowledge of the future inflow for each reservoir so that the optimization can be conducted to determine the optimal release. Another limitation is that the release can deviate from the actual value because of simplifications of the objective function and errors from irrigation demand calculation. The algorithm does not represent reservoirs with multi-year operational policies (*Adam et al.,* 2007) and also requires to run the model many times to optimize the reservoir release.

Adam et al. (2007) modified the *Haddeland et al.* (2006) reservoir model parameterization to include: 1) estimated minimum flow based on observed mean winter flow, 2) reservoir filling phase, 3) storage-area-depth relationship following the regular shape approximation of *Liebe et al.* (2005), 4) a seasonally varying hydropower production economic value that can be calibrated for hydropower production instead of a constant one. *van Beek et al.* (2011) further modified the retrospective inflow assumption to prospective model by approximating the upcoming operational year inflow based on previous years' inflow (requires historical inflow observation) and then adjusting the release and demand every month using the actual inflow as estimated from a hydrologic model.

Solander et al. (2015) tested and compared six generic equations to represent reservoir release and storage simulations. The complexity of equations tested varies from the simplest case that assumes that reservoir outflow equals inflow (no-reservoir assumption), to a more complex representation using separate linear functions during reservoir filling and release periods. While the reservoir filling and release seasons were identified using long-term mean temperature, their respective release equations are configured as a function of reservoir inflow, storage, and optimized seasonal empirical parameters. Their results for California reservoirs showed that the equation dependent on inflow is best for the recharge season, while release during the drawdown season was better represented as a function of storage. Despite failing for highly regulated reservoirs, their study demonstrated the possibility of generalizing the seasonal empirical parameters as a function of the ratio between winter inflows and storage capacity. However, further testing is required to examine the usefulness of *Solander et al.* (2015)'s methodology in different region, such as cold regions, with different filling and release seasonality.

Although the inflow-and-demand based models provide improved results compared to the natural lake approach, these models do not accurately reproduce observed flows (*Adam et al.*, 2007; *Haddeland et al.*, 2006; *Coerver et al.*, 2018). Overall, while the above methods have better flexibility for coupling with global hydrological and land surface models, the methods have limitations in accounting for details of reservoir operation. For an adequate representation of reservoirs, particularly multi-purpose reservoirs and/or those with multi-year carry-over capacity, it is important to consider *reservoir zoning* and adjust reservoir release formulations for different storage levels. The absence of this consideration may limit the capability of this group of methods in representing complex reservoir operations.

3.3.3 Neural network-based methods

Artificial neural network (NN) models have been applied to establish data-driven rules that relate reservoir storage, inflow and release data. This type of model requires (1) extensive data on reservoir release, storage, inflow, but minimal prior expert knowledge of the reservoir operation,

and (2) extensive training of a model for each individual reservoir to deduce the reservoir operation rules. Neural network techniques have been widely used beyond reservoir operation applications (e.g., flood forecasting, streamflow simulation, water quality (*Maier and Dandy*, 2001; *Razavi and Karamouz*, 2007)) and more recently have shown promise in reproducing historical reservoir operations (*Coerver et al.*, 2018).

The study of *Coerver et al.* (2018) provides detailed background on NN applications for deduction of reservoir operation rules, and also demonstrates the performance of NN-based fuzzy rules to describe the reservoir release decisions. The analysis of *Coerver et al.* (2018) involves different levels of input complexity for the neural network setups, such as the importance of accounting for inflow prediction and time of the season on the reservoir operation performance. Another similar application was shown by *Ehsani et al.* (2016) who demonstrated a general reservoir operation scheme that uses an NN technique to map the general input/output relationships to actual operating rules of seventeen dams. *Ehsani et al.* (2016) demonstrated the possibility of aggregating multiple reservoirs that are closely located, so that their integrated effect can be accounted for in large-scale hydrological modelling studies. In a subsequent study, *Ehsani et al.* (2017) integrated the reservoir model of *Ehsani et al.* (2016) into a global water security model to study reservoir operations under climate change.

While these studies demonstrated that, the NN-based models can reproduce historical reservoir operation data and possibly outperform the widely used reservoir simulation models such as those of *Hanasaki et al.* (2006) and *Wisser et al.* (2010), the user of such models may have to deal with a fundamental limitation, i.e., their "black-box" nature. This limits their ability to provide insight into the underlying mechanisms of reservoir operation, and might mask possible shortcomings in a derived NN model. Further, the credibility of their performance in extrapolation beyond the historical data can be in question, as they ignore the expert knowledge available on the actual physical and socio-economic processes that govern reservoir operations. Together, these limit the interpretation of results and their applicability in a changing environment. There have

been some recent research efforts to reformulate neural networks such that they can overcome these limitations (e.g., see *Razavi and Tolson*, 2011).

3.3.4 Target storage and release based methods

The target storage-and-release based methods aim to emulate actual rule curves (i.e., reservoir target storage and release for different times of the year) that guide reservoir operators to decide on downstream releases (*Burek et al.*, 2013; *Yates et al.*, 2005; *Neitsch et al.*, 2005). The target levels of storage divide the total reservoir storage capacity into multiple zones. For example, in the SWAT model (*Arnold et al.*, 1998), a reservoir model is available in which the total storage of a reservoir is divided into sediment, principal, flood control, and emergency flood control zones where zones are either specified by the user or as a function of soil moisture wetness (*Neitsch et al.*, 2005). *Wu and Chen* (2012) modified this approach by changing the reservoir zoning model and developed a reservoir release simulation strategy that uses a decision-based parameterization to better fit both storage and release of multi-purpose reservoirs. However, they reported only one application of this strategy to a local-scale reservoir, and its comprehensive evaluation needs to be performed on other reservoirs in other regions with different climates, levels of regulation, and allocation objectives.

Zhao et al. (2016) integrated a reservoir regulation module into a hydrology model, requiring user-specified (based on observed data) values to divide the reservoir into inactive, conservation and flood control zones. In their module, the release from the conservation zone is determined using water demand, which includes multi-sectorial demand and environmental flow. The release from the flood storage zone is decided as a function of inflow (classified as flood inflow or non-flood inflow), downstream channel current discharge and downstream maximum discharge. At a time of flood, if the downstream discharge is below the maximum limit, release from flood storage zone is estimated using available storage above conservation zone, multiplied by a weight parameter which allows to release more water. If the downstream discharge is at maximum capacity there is no release from flood storage zone. Finally, any storage above the flood

storage zone is automatically released. Additionally, *Zhao et al.* (2016) added the possibility to operate reservoirs conjunctively by giving release priorities to immediate downstream demands and by limiting the release if the downstream reservoir is within the flood storage zone. The results of the reservoir integration showed improved capability of the hydrological model to simulate water storage and release for Lake Whitney and Auilla Lake in Texas. The limitation of the *Zhao et al.* (2016) methodology for wider application is that there is no generic formulation of reservoir zoning (requires user specification) and evaluation was only performed on two reservoirs.

Similarly, *Burek et al.* (2013) divided the total reservoir storage into conservative, normal and flood zones within the LISFLOOD model, and defined releases in accordance with these storage zones using multiple-linear regression. *Zajac et al.* (2017) showed the applicability of this method to capture the effects of lakes and reservoirs globally using a parameterization that depends on naturalized inflow and maximum storage. Their results showed that the inclusion of reservoirs and lakes in a hydrologic model through this method helped improve streamflow simulation for many stations, but the performance in replicating observed storage dynamics was not reported.

Overall, the primary advantage of methods in this category is that they allow approximation of reservoir-release policy and have the potential of making use of detailed data on a reservoir when available. Their main limitation, however, is their relatively high data demands. When data are available, methods under this category have the potential to enhance the representation of dams and reservoirs in terms of both reservoir storage and release, while adapting to the seasonality and change in operations on different time scales from daily to seasonal. These methods seem advantageous compared to NN-based models as their functioning is transparent, accounting for the governing processes, while requiring similar data.

Given the advances in the field and the growing availability of data sources, the target storage-and-release methods seem to be the most promising, as they can better simulate the reservoir operation dynamics (the dynamics of both storage and release). The data requirement includes data on observed inflow, observed release, observed storage (level) and reservoir physical

characteristics. Reservoir level data are available for most lakes and reservoirs in the public domain, particularly in North America. These data can be converted to reservoir storage using reservoir elevation-area-volume relationships or by using area-volume relationships approximated by regular geometric shapes (*Yigzaw et al.,* 2018; *Liebe et al.,* 2005; *Lehner et al.,* 2011). Inflows to and releases from a reservoir can be approximated by streamflow stations located upstream and downstream of the reservoir, respectively. Further, satellite missions such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (*Savtchenko et al.,* 2004) and satellite radar altimetry are providing information on lake and reservoir surface area dynamics and reservoir water elevation for some large reservoirs. The combination of MODIS and satellite radar altimetry allows to derive storage-area-depth relationships (*Gao et al.,* 2012; *Andreadis et al.,* 2007; *Zhang et al.,* 2014; *Yoon and Beighley*, 2015). The planned Surface Water and Ocean Topography (SWOT) mission (2021) (*Garambois and Monnier*, 2015) will increase the availability of water level data for smaller rivers (with widths going down to 100 m) that can be potentially converted to discharge to estimate reservoir inflows and downstream reservoir releases.

3.4 Material and methods

This study aimed to develop an improved reservoir model that better emulates reservoir operation for large-scale hydrologic modelling application in terms of both reservoir storage and releases, following the previous advances in target-storage-and release-based methods reviewed in Section 3.3.4. In this section, we present the characteristics and formulation of our reservoir model. The reservoir water balance is maintained using the continuity equation, as shown in Eq. 1. The aim is to estimate an unknown storage S_t and release Q_t at the current time step based on the storage at the previous time step S_{t-1} and precipitation (P) over the reservoir, evaporation (E) from the reservoir and inflows (I) during the current time step. When integrated within an H-LSM model, the inflow will be the modelled value of the upstream catchment that would account for delays in the precipitation-runoff generation and routing. This equation is solved in conjunction with the parametrization equations presented in the next section for reservoir releases to compute S_t and Q_t.

$$\frac{S_t - S_{t-1}}{\Delta t} = \frac{I_t + I_{t-1}}{2} - \frac{Q_t + Q_{t-1}}{2} + \frac{P_t + P_{t-1}}{2} - \frac{E_t + E_{t-1}}{2} \quad (1)$$

3.4.1 Proposed reservoir operation model

A detailed description of our proposed target storage-and-release model (or target-release model for brevity) is provided here. This model is formulated in the form of parametric piecewise linear functions that approximate the reservoir release rules that may be used by reservoir operators. This model can be set up on any time scale; in the case studies reported here, we define the target levels to dynamically change over time. We call the model the "Dynamically Zoned Target Release (DZTR)" Model. Piecewise linear function-based reservoir operation models have already been used to solve complex reservoir operations and water resources management problems (e.g., *Razavi et al.,* 2013; *Asadzadeh et al.,* 2014). A systematic integration of such models into large-scale hydrological modelling has been reported in *Burek et al.* (2013) as implemented in the LISFLOOD hydrological model, and in Neitsch *et al.* (2005) as implemented in the SWAT model. Our DZTR model is a generalization of the method developed by *Razavi et al.* (2013), which may also be viewed as a modification to the model proposed by *Burek et al.* (2013) in terms of parametrization and reservoir zoning. **Fig. 3.1** shows the schematic representation of DZTR; **Fig. 3.1a** shows the reservoir zoning and **Fig. 3.1b** shows the piecewise-linear functions to estimate the release for each zone based on DZTR.

The DZTR model divides reservoir storage into five zones in a similar fashion to *Wu and Chen* (2012) and *Burek et al.* (2013), namely dead storage, critical storage, normal storage, flood storage and emergency storage. Whenever storage is below the emergency storage zone, release only occurs through the bottom outlet, but when the storage is within that zone, release happens through both a bottom outlet and the spillway. In the absence of data, the dead storage (Zone 0) is assumed to be 10% of the maximum storage, after *Döll et al.* (2009). To estimate the remaining storage zones in cases where no operational information on a reservoir is available, we propose two alternative strategies: (1) setting the zones based on some suggested exceedance probabilities from historical reservoir storage time series, (2) optimizing these zones to reproduce the observed

storage and release time series. Target releases for each zone can be obtained in a similar fashion. These target storages and releases are allowed to vary each month (or on any other arbitrarily selected time resolution) to allow a better representation of the seasonality of reservoir operation.

When reservoir storage is within the dead storage zone (Zone 0), the reservoir release is zero (Eq. 3). In Zone 1 (critical storage zone), the reservoir release is a function of storage at a given time step and the critical release target value (Eq. 4). In this zone, the reservoir operates to avoid storage depletion while trying to support environmental flow requirements defined as a critical (or minimum) release. In Zone 2 (normal storage zone), the reservoir release is purely governed by reservoir storage and varies between critical and normal release targets (Eq. 5). In this zone, the downstream release is greater for higher levels of storage. In Zone 3, the release decision considers both reservoir storage and inflow in that time step as well as the normal and maximum release targets (Eq. 6). When in this zone, two scenarios may occur: (A) the amount of inflow in a time step is equal to or less than the normal release rate; (B) the amount of inflow in this time step is greater than the normal release rate. As formulated in Eq. 6, in the case of scenario B, the inflow rate comes to play to augment the release in an attempt to keep the reservoir level within the normal storage zone. Scenario B is expected to occur more frequently in smaller reservoirs that only have "within-year" storage capacity, while scenario A should be more commonly seen with larger reservoirs that have "multi-year" carry over capacity. Hanasaki *et al.* (2006) suggested that reservoirs that have a ratio of storage capacity to mean annual inflow (referred to as c) of less than 0.5 be assumed as within-year reservoirs and the ones with a ratio of 0.5 and above be considered as multi-year reservoirs. Other values for this threshold were also suggested in the literature; e.g. *Wu and Chen* (2012) used a c value of 0.3. In this study, scenario A is used for reservoirs that have multi-year capacity (c>0.5) and scenario B for reservoirs that have within-a-year capacity (c>0.5). Lastly, in Zone 4 (emergency storage zone) the reservoir algorithm operates to avoid reservoir overtopping by releasing the larger of the maximum release target or all excess storage above the maximum storage value (the flood storage) constrained by the downstream channel capacity Q_{mc}. If not specified, a rough estimate of the downstream

channel capacity value could be the 99th percentile of non-exceedance probabilities of discharges from historical data.

Zone 0 $Q_t = 0$ $[S_t < 0.1S_{max}]$ (2)

Zone 1 $Q_t = \min\left(Q_{ci}, \frac{S_t - 0.1S_{max}}{\Delta t}\right)$ $[0.1S_{max} < S_t \leq S_{ci}]$ (3)

Zone 2 $Q_t = Q_{ci} + (Q_{ni} - Q_{ci})\frac{(S_t - S_{ci})}{(S_{ni} - S_{ci})}$ $[S_{ci} < S_t \leq S_{ni}]$ (4)

Zone 3A $Q_t = Q_{ni} + (Q_{mi} - Q_{ni})\frac{(S_t - S_{ni})}{(S_{mi} - S_{ni})}$ $[S_{ni} < S_t \leq S_{mi}]$ (5A)

Zone 3B $Q_t = Q_{ni} + \max\{(I_t - Q_{ni}), (Q_{mi} - Q_{ni})\}\frac{(S_t - S_{ni})}{(S_{mi} - S_{ni})}$ $[S_{ni} < S_t \leq S_{mi}]$ (5B)

Zone 4 $Q_t = \min([\max\left(\frac{(S_t - S_{mi})}{\Delta t}, Q_{mi}\right)], Q_{mc})$ $[S_{mi} < S_t]$ (6)

where I_t, Q_t and S_t are inflow, release and storage at time step t. S_{ci} , S_{ni} and S_{mi} are critical, normal and maximum storage targets for month i. Q_{ci} , Q_{ni} and Q_{mi} are critical, normal and maximum release targets for month i. Q_{mc} is maximum channel capacity parameter. The unit for inflow, release and target release parameters are in m^3 s^{-1}. The unit for storage and target storage parameters are in m^3.

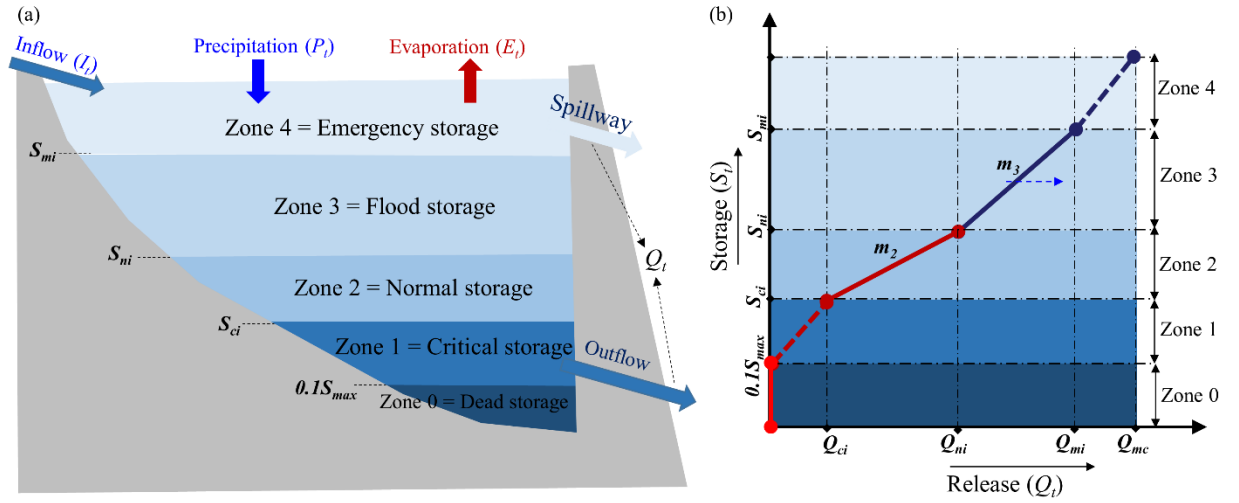


Figure 3. 1 The schematic representation of reservoir zoning and the storage-release function: (a) Four (active) reservoir zones with inflows and outflows; (b) piecewise linear reservoir release function, m_1, and m_2 control the slope of the release curve and they change monthly. The upward blue arrow is to indicate that inflow to the reservoir may also be considered in determining the release in zone 3

3.4.2 Evaluation criteria

We evaluated the performance of the proposed reservoir operation model in emulating the outflow and storage data collected for many reservoirs around the world. As this model was intended to be integrated into large-scale H-LSMs, we further evaluated it when embedded in the MESH model (Modélisation Environmentale–Surface et Hydrologie) (*Pietroniro et al.*, 2007). For all of these evaluations, we used Nash-Sutcliffe Efficiency (NSE) (*Nash and Sutcliffe*, 1970) and Kling-Gupta Efficiency (KGE) (*Gupta et al.*, 2009) as the metrics to assess the goodness-of-fit of the model to observed reservoir outflow and storage data.

3.4.3 Identification of reservoir operation model parameters

As demonstrated in Section 3.4.1, the proposed reservoir operation model has six parameters ($S_{ci}, S_{ni}, S_{mi}, Q_{ci}, Q_{ni}$ and Q_{mi}) that can vary for different times of the year. We recommend varying these parameters on a monthly basis, while other time resolutions are also possible. To normalize the parameters and their ranges across different types and sizes of reservoirs, for every

reservoir, we use cumulative distribution functions (CDFs) of historical storage and release values; see **Fig. 3.2** CDFs of the Lake Diefenbaker reservoir (Gardiner dam) in the Saskatchewan River Basin, Canada as an example. Our preliminary analysis indicated that target storage and release values corresponding to 10%, 45%, and 85% non-exceedance probabilities generally perform reasonably well. We call these our 'generalized parameterization'.

However, optimal values of parameters for a given reservoir can be identified, when data are available, through optimization and parameter identification techniques (*Maier et al.*, 2019; *Guillaume et al.*, 2019). For this purpose, we use a bi-objective optimization approach, that follows, that begins with the generalized parameter values as the starting point and optimizes the model fit to both storage and release data simultaneously.

$$\underset{x\in\Omega}{\text{maximize}} \qquad F(x) = (f_1(x), f_2(x)) \tag{7}$$

where x is a vector of decision variables (parameter values), Ω is the decision space, $f_1(x)$ is NSE(flow) measuring the goodness-of-fit in reproducing observed release, and $f_2(x)$ is NSE(storage) measuring the goodness-of-fit in reproducing observed storage dynamics.

For parameter identification on a monthly basis, a total of 72 decision variables were used in the optimization. We chose rather arbitrarily the storage and release target intervals that correspond to [5-35%], [35-75%], and [75-95%] non-exceedance probabilities as the ranges of variation for critical, normal, and maximum (flood) storage and release, respectively.

The bi-objective optimization problem to calibrate 72 reservoir target release and storage parameters was conducted using a multi-algorithm, genetically adaptive multi-objective method (AMALGAM) optimization algorithm (*Vrugt and Robinson*, 2007). AMALGAM was selected because it provides effective and reliable solutions for multi-objective optimization using multiple search operators (genetic algorithm, particle swarm optimization, adaptive metropolis search, and differential evolution) and self-adaptive offspring creation. *Vrugt et al.* (2009), *Wöhling and Vrugt* (2011), *Zhang et al*. (2011), *Raad et al.* (2009), *Dane et al.* (2010) and others showed that the performance of AMALGAM model parameter calibration was better than, or equivalent to, some

other calibration algorithms across different complex response surfaces. AMALGAM was run using an initial population size of 100, resulting in a total of 15,000 model evaluations to estimate final Pareto solutions for every single reservoir.

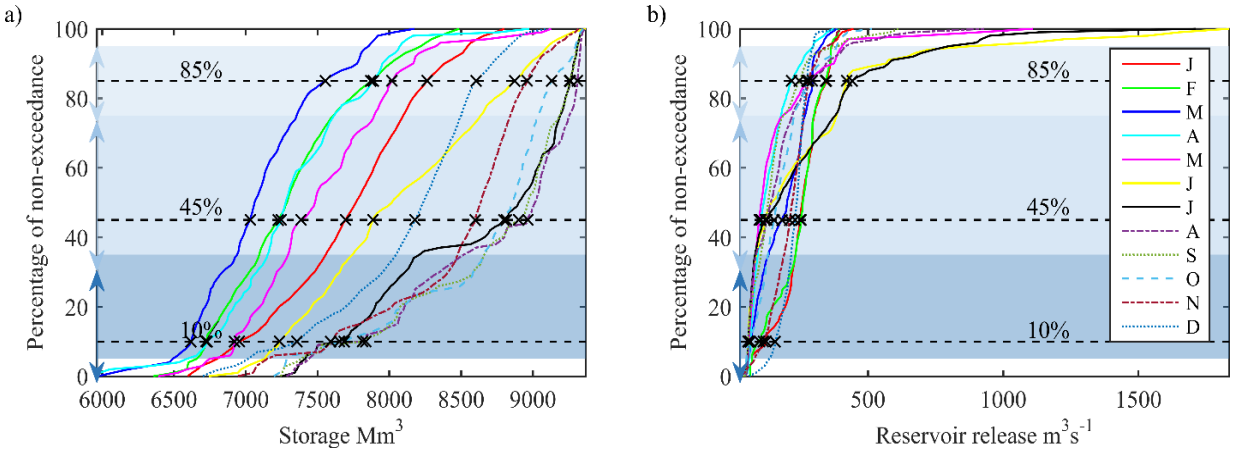


Figure 3. 2 Monthly Cumulative Distribution Function (CDFs) (a) Storage CDFs of Gardiner dam (b) Reservoir release CDFs of Gardiner dam. Double arrows on *y*-axis shows parametrizations ranges for each generalized parameters.

3.4.4 Comparison of reservoir operation models

We compared the performance of our DZTR model against those of *Hanasaki et al.* (2006) and *Wisser et al.* (2010) using NSE and KGE performance metrics defined on both storage and release simulations. The comparisons were made only for selected non-irrigation reservoirs because their irrigation reservoir formulation requires additional data on water demands. For the method of *Wisser et al.* (2010), reservoir release was estimated under two conditions as shown in Eq. 8 and storage dynamics are estimated from the reservoir water balance (Eq. 9).

$$Q_t = \begin{cases} \kappa I_t & I_t \geq I_m \\ \lambda I_t + (I_m - I_t) & I_t < I_m \end{cases} \quad (8)$$

$$S_t = S_{t-1} + (I_t - Q_t)\Delta t \quad (9)$$

where κ and λ are empirical constants set to 0.16 and 0.6, respectively and I_m is the mean annual inflow ($m^3\ s^{-1}$) and I_t is inflow to the reservoir ($m^3\ s^{-1}$) at time *t*.

In the method of *Hanasaki et al.* (2006), the release from non-irrigation reservoirs was estimated by multiplying the mean annual inflow by release constraining coefficients (Eq. 10). The release constraining coefficient for every given operation year were estimated by dividing the initial storage of that year by the maximum storage and the value is calculated for each simulation year (Eq. 11). The start of the operational year was considered to be the month when the mean monthly inflow shifts from being greater to being lower than the mean annual inflow.

$$r_{m,y} = \begin{cases} k_{rls,y} \cdot r'_{m,y} & (c \geq 0.5) \\ \left(\frac{c}{0.5}\right)^2 \cdot k_{rls,y} \cdot r'_{m,y} + \left(1 - \left(\frac{c}{0.5}\right)^2\right) \cdot i_{m,y} & (0 \leq c < 0.5) \end{cases} \quad (10)$$

$$k_{rls,y} = \frac{S_{first,y}}{\alpha \cdot S_{max}} \quad (11)$$

where c is the ratio of maximum reservoir storage to the mean total annual inflows; and $k_{rls,y}$ is the release coefficient; $r'_{m,y}$ is the provisional monthly release ($m^3\ s^{-1}$) which is equal to mean annual inflow ($m^3\ s^{-1}$); α is a dimensionless constant set to 0.85. Eq. 10 differentiates between multi-year and single year storage reservoirs based on a threshold value of 0.5 for *c*.

3.4.5 MESH modelling system

MESH is Environment and Climate Change Canada's Land Hydrology-Land Surface Modelling System (Pietroniro *et al.*, 2007) and has been widely used in different parts of Canada (*Davison et al.*, 2016; *Haghnegahdar et al.*, 2017; *Yassin et al.*, 2017; *Sapriza-Azuri et al.*, 2018; *Berry et al.*, 2017a, 2017b, 2017c). MESH is a grid-based modelling system composed of three components: (1) the Canadian Land Surface Scheme (CLASS) (*Verseghy*, 1991; *Verseghy et al.*, 1993), (2) lateral movement of surface (overland) runoff and sub-surface water (interflow) to the channel system within a grid cell and (3) hydrological flow routing using WATROUTE from the WATFLOOD hydrological model (*Kouwen et al.*, 1993).

Currently, the reservoir representation in MESH model is rudimentary. MESH offers two approaches to account for reservoir operations. In the first approach, the observed reservoir release rate at the reservoir location is provided as input to the model. In this approach, the flow from the catchment upstream of the reservoir is discarded as the release is replaced by observations, a process referred to as "*streamflow insertion*", which limits the utility of the model to simulate future scenarios for which releases are not yet known. This approach violates the water conservation law in the model and also creates discontinuities within the model setup, especially if there are reservoir cascades. Nevertheless, streamflow insertion could be used when coupling water management models with MESH, and these coupled models could be used to formulate scenarios for reservoir operations. As mentioned in the objectives, however, model coupling is not the focus of this study as we are looking to examine the internal representation of reservoir operations within CMs and LSMs. The second approach is a natural lake or uncontrolled reservoir representation model similar to that of *Döll et al.* (2003), which was shown to be unsuitable for highly managed reservoirs. To improve the reservoir representation in MESH, this study aims to incorporate the DZTR model for controlled reservoirs into the MESH framework and evaluate its performance.

3.4.6 Case studies and data

The dataset required to build and evaluate a reservoir operation model includes (1) reservoir physical characteristics such as the volume-level-area relationship and maximum capacity, which are static (in the absence of sedimentation or dam heightening), (2) time series of hydrologic variables such as inflow, release, and water level (or storage), and (3) environmental flows. In this study, we assembled such a dataset for 37 reservoirs located in several regions across the globe (**Fig. 3.3**) to test the model. These dams represent a wide range of storage sizes, from 0.132×10^9 m^3 to 162×10^9 m^3, spanning multiple orders of magnitudes. Most of these are located in the Western United States and Western Canada; while some are located in Vietnam, central Asian countries and Egypt. **Table 3.1** provides a summary of reservoir locations, construction years, main purposes, data periods, and other dam characteristics. Measured inflow, release, and storage

time series were collected from different sources. For reservoirs located in Canada, the data were acquired from Water Survey of Canada, Alberta Environment and Parks, and the Saskatchewan Water Security Agency. Data for the High Aswan dam were acquired from the Nile Basin Encyclopaedia via the Nile Basin Initiative. The data for other reservoirs were provided by the authors of previous studies (*Hanasaki et al.*, 2006 and *Coerver et al.*, 2018). Additional information about the degree of regulation, dam height, and catchment area were obtained from the GRanD database (*Lehner et al.*, 2011). Reservoir operation simulations were performed on daily and monthly bases with simulation periods varying from 8 to 62 years. The choice of simulation period and time scale was based on data availability (**Table 3.1**). The first year of the reservoir simulations was used for spin-up, while the first half of the remaining data periods were used for calibration and the second half for model validation.

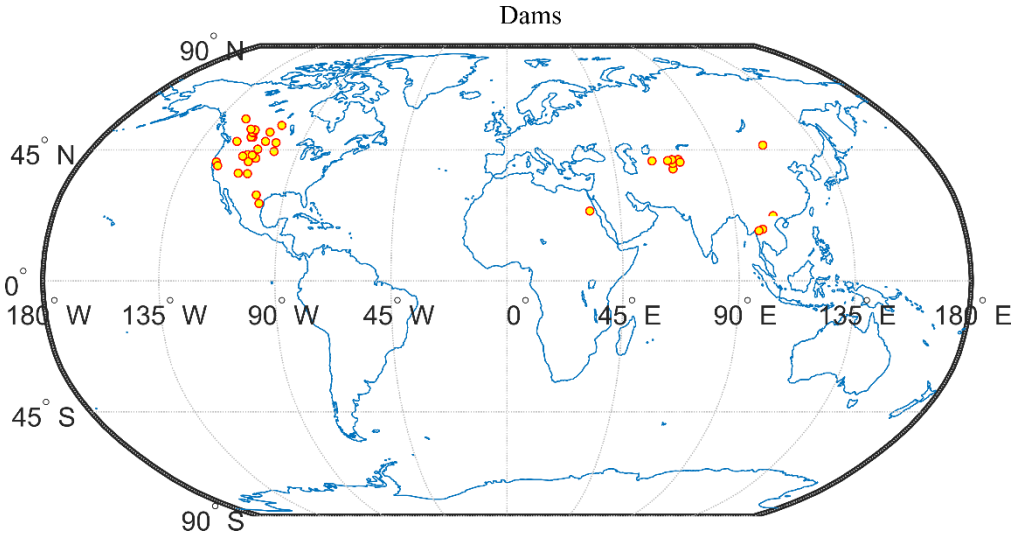


Figure 3. 3 Locations of dams used to evaluate the reservoir routing model.

We also evaluated the integration of our reservoir model into the MESH model on seven reservoirs in two major basins in Western Canada. Six of the test reservoirs (Gardiner, St Mary, Waterton, Oldman, Ghost and Dickson dams) are located within the heavily regulated Saskatchewan River basin (SaskRB) and one reservoir (Bennet dam) is located in the Mackenzie River Basin (MRB). For both of the basins, the MESH model was set up on a grid resolution of

0.125° and the data required to build the MESH model were obtained from different sources. The topographic data are based on the Canadian Digital Elevation Data (CDED) at a scale of 1:250,000 and were obtained from the GeoBase website (http://www.geobase.ca/). The data on seven climate forcing variables were obtained from Global Environmental Multi-scale (GEM) NWP model (*Côté et al.*, 1998) and Canadian Precipitation Analysis (CaPA) (*Mahfouf, et al.*, 2007). The land cover data used are based on a 2005 land-cover map from the Canada Centre for Remote Sensing (CCRS). Soil texture data were obtained from Soil Landscapes of Canada (SLC) data of Agriculture and Agri-Food Canada. The MESH parameter values were taken from previous studies for calibration to streamflow at major subbasins of the SaskRB and MRB.

Table 3. 1 Summary of reservoirs

Dam name	Country	Year	Main Purpose*	LONG (°)	LAT (°)	Dam height (m)	Capacity (MCM)	$c = \left(\frac{Capacity}{MAI}\right)$	Simulation Period (years)	Percentage Bias (PBIAS) Release vs Inflow
American Falls	USA	1977	IR	-112.87	42.78	32	2061.5	0.303	1978-1995 (18)	-3.29
Andijan[1]	Uzbekistan	1974	HP	73.06	40.77	115	1900	0.444	2001-2013 [m] (13)	-0.98
Bhumibol	Thailand	1964	IR	98.97	17.24	154	13462	2.645	1980-1996 (16)	-10.29
Big Horn	Canada	1972	HP	-116.32	52.31	150	1770	0.747	2002-2011 (10)	16.08
Bull Lake[1]	USA	1938	IR	-109.04	43.21	24	187.2	0.883	2001-2013(13)	-3.74
Canyon Ferry[1]	USA	1954	HP	−111.73	46.65	69	2464.4	0.543	1971-2011 [m] (40)	-1.46
Chardara	Kazakhstan	1968	IR	67.96	41.24	27	6700	0.354	2001-2010 [m] (10)	7.57
Charvak[1]	Uzbekistan	1977	HP	69.97	41.62	168	2000	0.284	2001-2010 [m] (10)	1.6
Dickson	Canada	1983	WS	-114.21	52.05	40	203	0.167	2005-2011(6)	27.3
E.B. Campbell[1]	Canada	1963	HP	-103.40	53.66	34	2200	0.153	2000-2011(12)	-1.69
Flaming Gorge[1]	USA	1964	WS	-109.42	40.91	153	4336.3	2.460	1971-2017(46)	-6.37
Fort Peck[1]	USA	1957	FC	-106.41	48.00	78	23560	2.210	1970-1999 [m] (30)	6.33
Fort Randal	USA	1953	FC	-98.55	43.06	50	6683	0.240	1970-1999[m] (30)	-1.43
Gardiner	Canada	1968	IR	-106.86	51.27	69	9870	1.460	1980-2011(32)	-3.44
Garrison[1]	USA	1953	FC	-101.43	47.50	64	30220	1.436	1970-1999 [m] (30)	-5.79
Ghost	Canada	1929	HP	-114.70	51.21	42	132	0.048	1990-2011(22)	5.43
Glen Canyon[1]	USA	1966	HP	-111.48	36.94	216	25070	2.230	1980-1996(17)	-6.87
Grand Coulee	USA	1942	IR	-118.98	47.95	168	6395.6	0.124	1978-1990(12)	-3.37
High Aswan	Egypt	1970	IR	32.88	23.96	111	162000	2.843	1971-1997 [m] (26)	-3.34
Int. Amistad	USA/Mexico	1969	IR	-101.05	29.45	87	6330	2.457	1977-2002(25)	-20.28
Int. Falcon Lake	USA/Mexico	1954	FC	-99.17	26.56	53	3920	1.045	1958-2001(43)	-14.48
Kayrakkum[1]	Tajikistan	1959	HP	69.82	40.28	32	4160	0.199	2001-2010 [m] (10)	1.19

Navajo	USA	1963	IR	-107.60	36.80	123	1278	1.744	1971-2011(40)	-21.07
Nurek	Tajikistan	1980	IR	69.35	38.37	300	10500	0.540	2001-2010 [m] (10)	0.28
Oahe Dam[1]	USA	1966	FC	-100.40	44.45	75	29110	1.244	1970-1999 [m] (30)	-5.366
Oldman River	Canada	1991	IR	-113.90	49.56	76	490	0.446	1996-2011(16)	3.98
Oroville[1]	USA	1968	FC	-121.48	39.54	235	4366.5	0.804	1995-2004(11)	4.20
Palisades	USA	1957	IR	-111.20	43.33	82	1480.2	0.242	1970-2000(31)	0.48
Seminoe	USA	1939	IR	-106.91	42.16	90	1254.8	1.048	1951-2013 [m] (63)	-4.10
Sirikit	Thailand	1974	IR	100.55	17.76	114	9510	1.834	1981-1996(16)	-7.32
St. Mary	Canada	1951	IR	-113.12	49.36	62	394.7	0.492	2000-2011(12)	0.16
Toktogul[1]	Kyrgyzstan	1978	HP	72.65	41.68	215	19500	1.393	2001-2010 [m] (10)	-6.34
Trinity	USA	1962	IR	-122.76	40.80	164	2633.5	1.470	1970-2000(31)	-4.18
Tyuyamuyun	Turkmenistan	N/A	IR	61.40	41.21	N/A	6100	0.204	2001-2010 [m] (10)	-2.43
W.A.C. Bennett	Canada	1967	HP	-122.20	56.02	183	74300	1.200	2003-2011(9)	5.41
Waterton	Canada	1992	IR	-113.67	49.32	55	172.7	0.258	2000-2011(12)	-10.34
Yellowtail	USA	1967	IR	-107.95	45.30	160	1760.6	0.489	1970-2000[m] (31)	-1.693

[*]Main purpose: WS-Water Supply, HP-Hydropower IR-Irrigation FC-Flood Control

[m] Represents monthly data and simulation

[1] Represents multiple reservoir models are compared on this reservoir

N/A – Not Available,

MCM – Million Cubic Meters.

3.5 Results and discussion

3.5.1 Evaluation of the Dynamically Zoned Target Release (DZTR) model with generalized parameters

Individual reservoir simulations were conducted by the DZTR model with generalized monthly storage and release parameter values set at non-exceedance probabilities recommended in Section 3.3 for representing the reservoir storage zones and their respective target releases. The evaluation of the DZTR model was based on the performance metrics and a comparison with the other reservoir operation approaches and a base case where the existence of a reservoir was ignored in a model, referred to as the "no-reservoir assumption". Under the no-reservoir assumption, the release was considered equal to inflow, and storage was considered constant, and as such, the performance metrics were computed by directly comparing inflow with observed release.

Fig. 3.4 shows performance metrics results of the DZTR model in terms of NSE and KGE for storage and release simulations compared to those of the base case. As shown in **Fig. 3.4a**, both

NSE (flow) and NSE (storage) results are greater than 0.25 and 0.5 for 90% and 50% of reservoirs, respectively. Although NSE (flow) results are greater than zero for all reservoirs, 1 of the reservoirs resulted in a negative NSE (storage) value. The no-reservoir assumption resulted in NSE (base-case) values of greater than 0.25 and 0.5 for 45% and 30% of reservoirs, respectively, which, in general, are much lower than those of the DZTR model. Under the no-reservoir assumption, 48% of the reservoirs resulted in a negative NSE (base-case). Almost all positive NSE (base-case) results were observed on reservoirs with $c<0.5$ such as Dickson, E.B. Campbell, Kayrakkum, Oldman and Tyuyamuyun (as explained in Section 3.4, c is the ratio of storage capacity to annual inflow volume). However, for reservoirs with $c>0.5$ such as Bhumibol, Flaming Gorge, Fort Peck, High Aswan, W.A.C. Bennett, the NSE (base-case) is negative, which indicates the significant influence of their regulations on the hydrograph shape. Similarly, **Fig. 3.4b** shows the evaluation of the different reservoir models based on the KGE metric (*Gupta et al.*, 2009). The values of KGE (flow) and KGE (storage) are greater than 0.25 and 0.5 for 100% and 86% of the reservoirs, respectively. The KGE (base-case) values of 21% of reservoirs are less than 0, while those of 57% and 49% of the reservoirs are greater than 0.25 and 0.5, respectively. The NSE and KGE results show that the DZTR with the generalized parameter values is capable of simulating flow and storage simulation well.

.

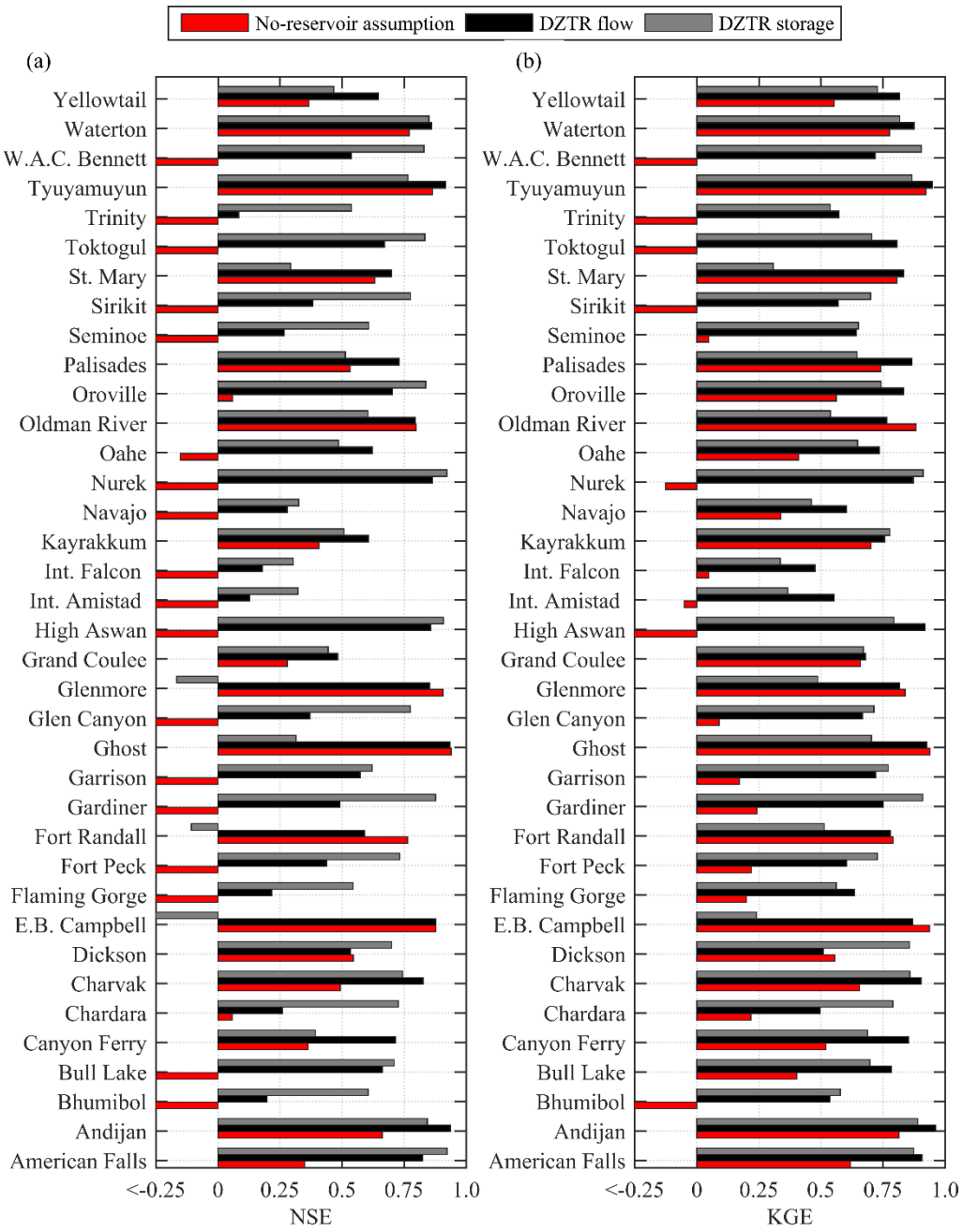


Figure 3. 4 Performance evaluation result of the DZTR model reservoir operation algorithm (a) NSE performance metrics, (b) KGE performance metrics.

Fig. 3.5 shows scatter plots between NSE, KGE, and the regulation level represented by *c*. The orientation of the scatter plot between NSE and KGE on flow and storage shows a strong positive correlation between the evaluation metrics, which indicates that both metrics provide somewhat similar evaluation information. **Fig. 3.5a** and **3.5b** show that both no-reservoir assumption and DZTR estimate the release more accurately for lower levels of regulation. As expected, the reduction of performance was pronounced for the no-reservoir assumption as the regulation level increased, while DZTR performance diminished by a much smaller extent (still positive values). Almost all low regulation level reservoirs (c<0.5) showed positive performance metrics which means the reservoir regulation does not strongly modify the flow regime, whereas the opposite case is true for highly regulated reservoirs (c>0.5) in which the reservoir regulation strongly changes the reservoir release. *Coerver et al.* (2018) also noted that low regulation level reservoirs are more dependent on the current time step inflow knowledge because their smaller influence on the flow regime. The method of *Hanasaki et al.* (2006) also recognizes the strong dependence of c<0.5 reservoirs on inflow to determine the release by configuring the release as a function of monthly mean inflow. Conversely, the relationship between the regulation level and the storage simulation performance (in terms of both KGE (storage) and NSE (storage)) did not show a strong correlation (**Fig. 3.5c**).

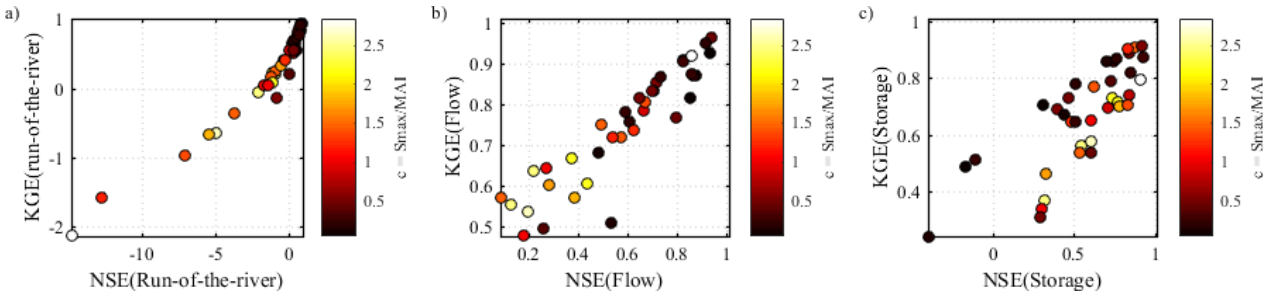


Figure 3. 5 Scatter plot between NSE and KGE with regulation scale represented in terms of *c* (a) NSE and KGE on no reservoir condition, (b) NSE and KGE on DZTR release, and (c) NSE and KGE on DZTR storage.

Fig. 3.6 compares the reservoir simulation and observation time series for the whole simulation period, while **Fig. 3.7** shows the long-term average of these simulations. Inflows are also included in **Fig. 3.6** and **Fig. 3.7** to show the regulation pattern and changes caused by reservoir operation. Both figures indicate that the DZTR model captures both release and storage dynamics well, reproducing the daily and monthly seasonality as well as the magnitude and timing of storage and releases for almost all reservoirs, especially for reservoirs with high regulation (multipurpose, multiyear reservoirs) such as American Falls, Bhumibol, High Aswan, Sirikit, Trinity, and W.A.C. Bennett dams. However, the simulations also show some systematic over- and under-estimations; for example, the simulations of Bhumibol, Fort Peck, High Aswan, Int. Falcon, Navajo, Bennet, and Int. Amistad reservoirs show continuous underestimation and overestimation of reservoir storage. Some reservoirs such as Trinity, Palisades, Kayrakkum, Flaming Gorge, and Garrison show underestimation and overestimation of reservoir storage only for some seasons. A closer look at American Falls, Flaming Gorge, Fort Peck, Glen Canyon, Navajo dams in **Fig. 3.6** indicates that the DZTR model reliably captured storage and release seasonality, inter-annual trends, and release pattern shifts during consecutive wet years 1982-1986 followed by consecutive dry years 1987-1993. Similar patterns can be observed for the Gardiner dam with good simulation results during both dry years (1984-1986, 1988-1989, 1999-2004) and wet years (1993, 2005, 2010-2011). Furthermore, as expected, **Fig. 3.7** shows that lowly-regulated reservoirs (c <0.5) have less impact on the flow regime, but with fairly significant storage seasonality (Oldman, E.B. Campbell, Palisades, Andijan). In general, the DZTR model with the generalized parameterization of reservoir zones and releases showed an improved performance and can be applied to any hydrological model (CM or H-LSM) that involves reservoir simulation.

It is important to note that for the case of a cascade of reservoirs, the parametrization of the DZTR model implicitly accounts, to some extent, for the upstream regulation effects by the upstream cascade reservoirs. This is because the regulated inflow is used for parametrizing downstream reservoirs, which reflects the regulation information of upstream reservoirs in the cascade. In reality, the operation of some cascade reservoirs are highly interlinked, particularly

during the flood season. The decision regarding the release from one reservoir accounts for the (forecasted) state of other reservoirs. Such dual- or multi-linked operation is however not accurately accounted for in the presented algorithm, because it assumes that each reservoir operates using its own storage state, inflow and target storage and releases. Such systems require detailed modelling of operations that is not usually attainable in large scale hydrological models. Depending on the purpose of the model, the modeller may decide to lump those reservoirs together to improve simulations downstream as in e.g., *Ehsani et al.* (2016).

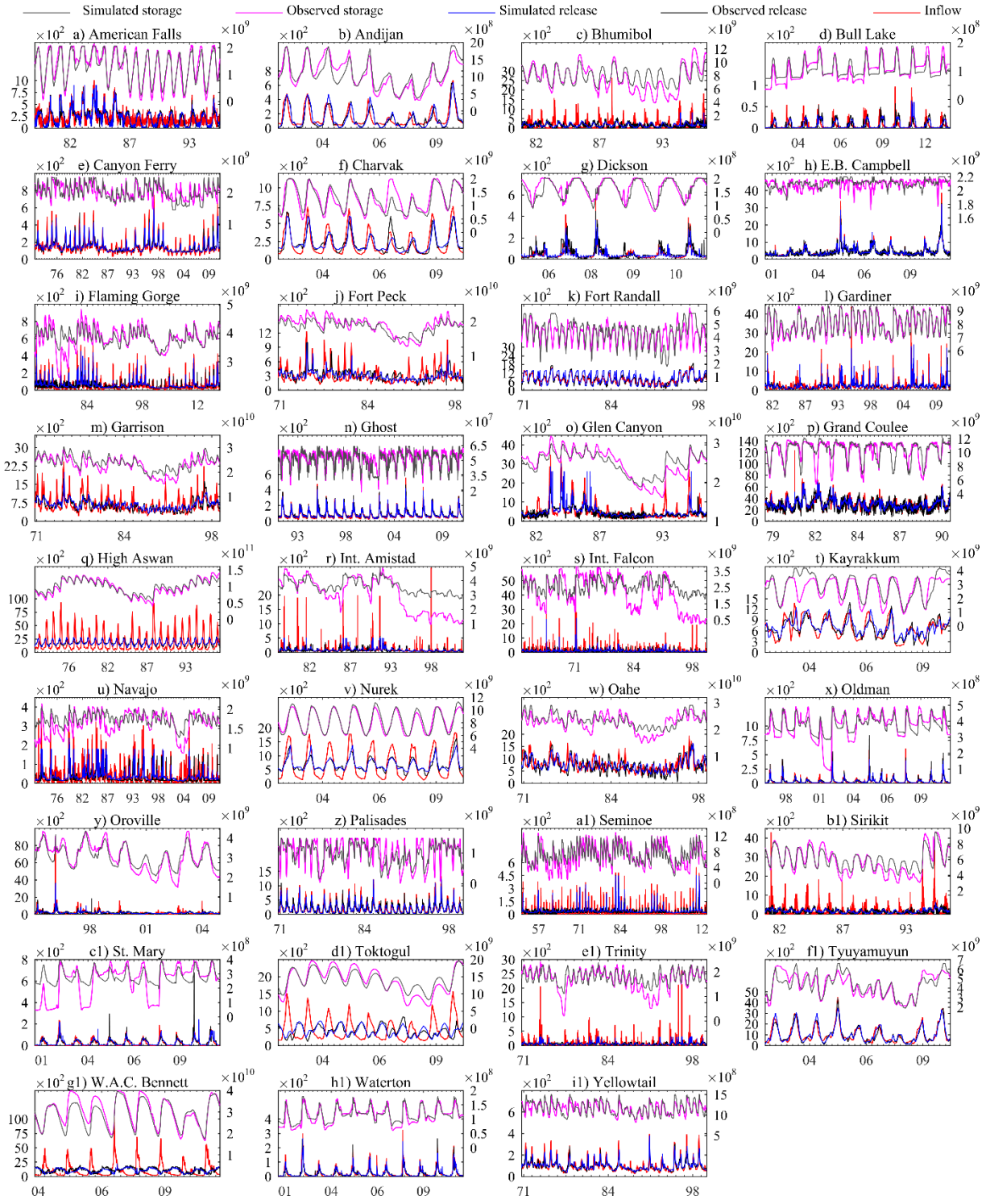

Simulated storage
Observed storage
Simulated release
Observed release
Inflow
a) American Falls
b) Andijan
c) Bhumibol
d) Bull Lake
e) Canyon Ferry
f) Charvak
g) Dickson
h) E.B. Campbell
i) Flaming Gorge
j) Fort Peck
k) Fort Randall
l) Gardiner
m) Garrison
n) Ghost
o) Glen Canyon
p) Grand Coulee
q) High Aswan
r) Int. Amistad
s) Int. Falcon
t) Kayrakkum
u) Navajo
v) Nurek
w) Oahe
x) Oldman
y) Oroville
z) Palisades
a1) Seminoe
b1) Sirikit
c1) St. Mary
d1) Toktogul
e1) Trinity
f1) Tyuyamuyun
g1) W.A.C. Bennett
h1) Waterton
i1) Yellowtail

Figure 3. 6 Daily and monthly reservoir simulations using the DZTR model with a generalized parametrization, the *x* axes show month/year, the primary *y* axes show release (m^3 s^{-1}) and the secondary *y* axes show storage (m^3).

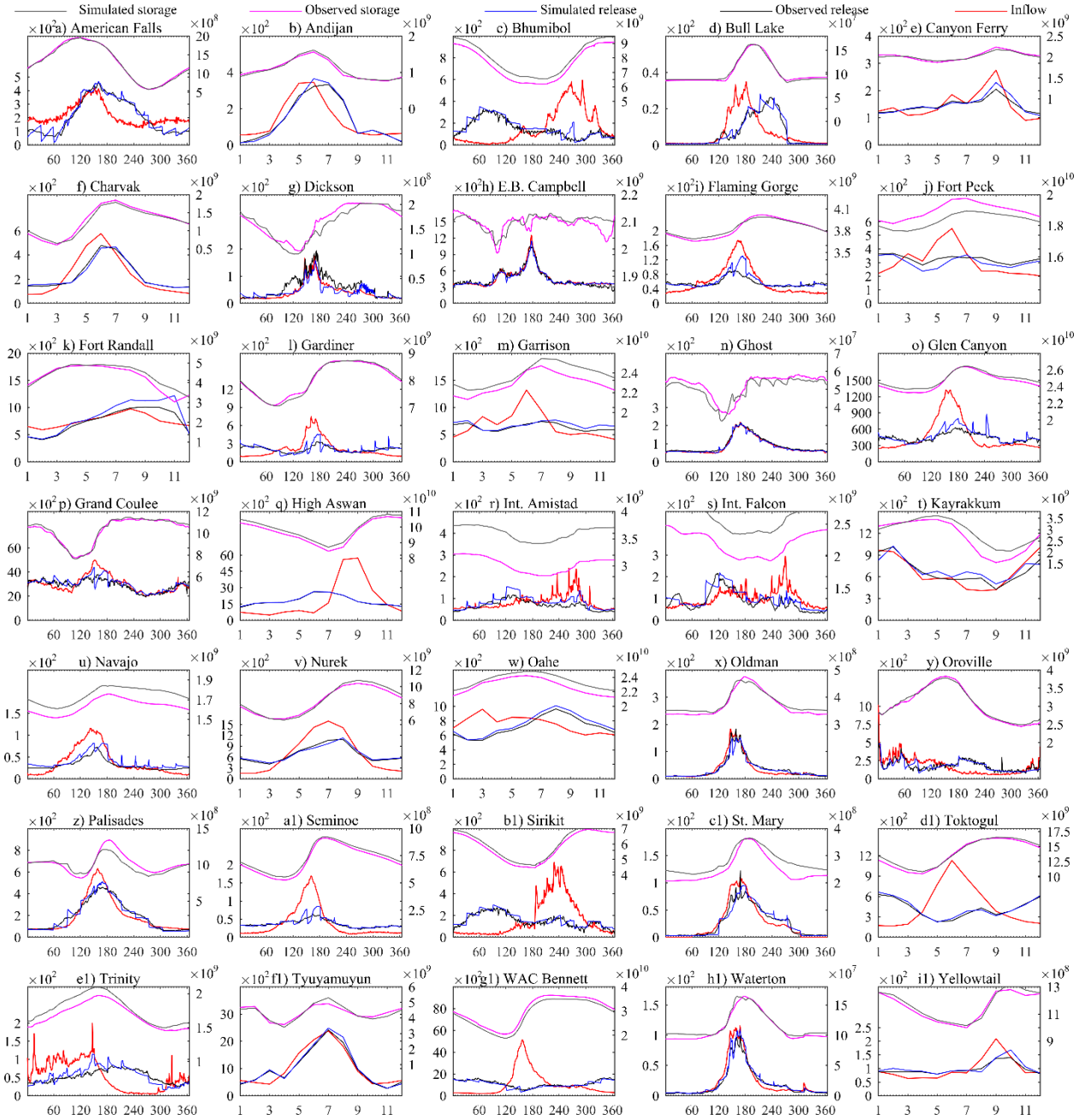


Figure 3. 7 Long-term average daily or monthly reservoir simulations with generalized parametrization, the *x* axes show days (1-365) or months, (1-12) the primary *y* axes show release (m^3 s^{-1}) and the secondary *y* axes show storage (m^3).

3.5.2 Comparison with previously developed reservoir operation models

To further illustrate the reliability of the DZTR model in simulating reservoir operations, a comparison with the methods of *Hanasaki et al.* (2006) and of *Wisser et al.* (2010) was conducted as shown in **Fig. 3.8**. The comparison shows that the DZTR model provides a considerable improvement according to all of the performance criteria, notably NSE (Storage) and NSE (Flow), except in the case of the E.B. Campbell dam where *Hanasaki et al.*'s method showed similar performance to DZTR. The method of *Hanasaki et al.* (2006) outperformed that of *Wisser et al.* (2010). Out of the 13 reservoirs compared, the DZTR resulted in positive values for both NSE (Storage) and NSE (Flow) for all except for E.B. Campbell storage. The method of *Hanasaki et al.* (2006) and *Wisser et al.* (2010) resulted in eight and five reservoirs with positive NSE (Flow) respectively (**Fig. 3.8a**), while both produced negative values for NSE (storage) for all the reservoirs compared (**Fig. 3.8c**). A similar performance pattern was observed for KGE metrics for flow and storage. In addition, we compared the DZTR result shown in **Fig. 3.7** and **3.8** with the results reported in *Coerver et al.* (2018) who applied a fuzzy-neural network model to extract 11 operating rules. This comparison showed that the performance of our generalized parameterization is comparable to that of *Coerver et al.* (2018) in simulating reservoir release; note that performance on storage is not reported in *Coerver et al.* (2018). This indicates that the simple parameterization applied in the DZTR model can provide a solution that is at least as effective as that of a neural network-based model. Equally importantly, the DZTR model is transparent, as opposed to neural network methods that are often criticized as being a "black box".

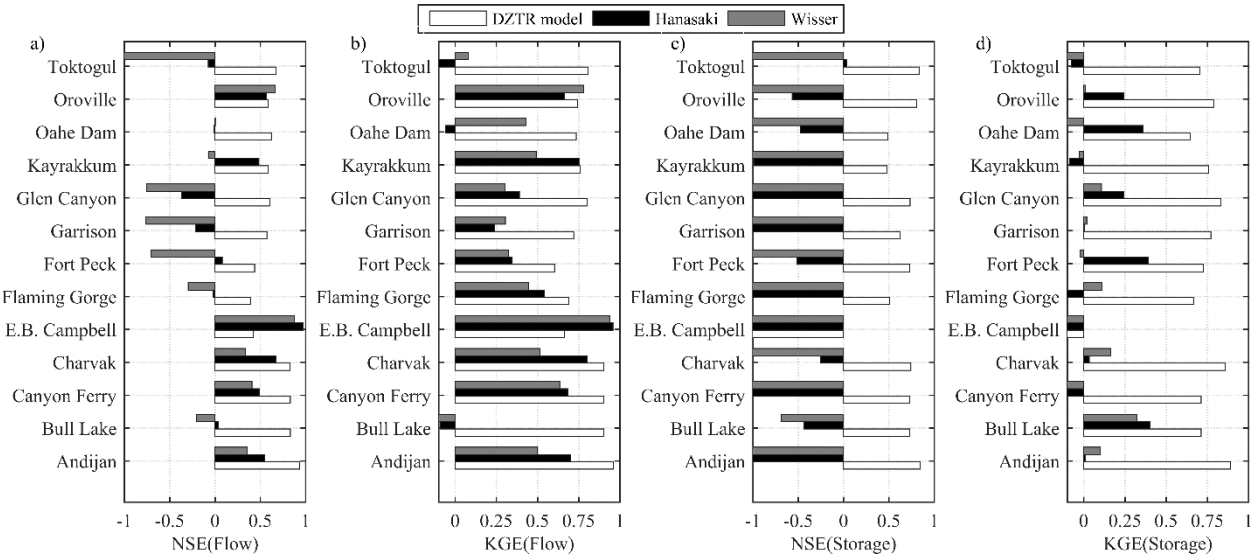


Figure 3. 8 A comparison of our proposed reservoir operation model with generalized parameters with the models of Hanasaki *et al.* (2006) and Wisser *et al.* (2010).(a) NSE(Flow), (b) KGE(Flow), (c) NSE(Storage), (d) KGE(Storage).

The above comparisons were conducted for non-irrigation reservoirs because water demand data are needed to use the *Hanasaki et al.* (2006) method for irrigation reservoirs. In the case of the DZTR approach the idea is that the DZTR model operates in such a way to infer existing operational rules that cater for those demands. Thus, the release from DZTR accounts, implicitly, for downstream demands as per the intended purpose of the reservoir whether it is for flood control, irrigation, hydropower, etc. or any combination of these. The case study dams include reservoirs with different purposes as shown in Table 3.1. The DZTR approach showed good performance for reservoirs with different purposes.

If the reservoir purpose is irrigation, the target releases from DZTR are to satisfy irrigation demands because the parameterization is optimized based on observed releases. The release from an irrigation dam will be available for abstraction at the predefined abstraction points downstream of the dam. The abstraction and distribution can be implemented as separate modules as done within the MESH land surface model (*Yassin et al.,* 2019). In such an implementation, MESH takes care of (1) calculation of actual irrigation demand for a configured irrigation area, (2) water

abstraction from defined abstraction point along the river below the dam and (3) distribution across the irrigation fields. Regarding the return flow, the excess water flows from the irrigation areas are assumed to join the nearest stream within the model grid cell.

The DZTR model can in principle handle multi-purpose reservoirs, e.g., a reservoir that is used simultaneously for hydropower generation, irrigation water supply, and flood control (e.g., the High Aswan Dam in Egypt which is one of the studied reservoirs), the DZTR provides the release based on the inflow, and storage conditions and that will be available for irrigation downstream. Hydropower does not consume water but returns it back to the river (except in rare cases where it returns to a different channel). Flood control is directly accounted for in the scheme and becomes relevant when storage is within the flood storage zone. Further, the flexible formulation of DZTR allows to implicitly change the priorities in operation for selected time periods (e.g., months or seasons) by changing the target storage values during flood periods (e.g., the storage target before the onset of snowmelt). During these flood months, lowering the target storage would increase the buffer for flood control. Conversely increasing the target storage during other months would be desirable to store water and release during irrigation months. When the scheme is optimized using inflow, release, and storage data, the parameterizations capture these priorities implicitly as expressed in the data. When inflow data are lacking, the generalized parametrization will set the storage zones based on the suggested exceedance probabilities (that were deduced based on all reservoirs used in the study) and the priorities can be assumed as pre-defined.

3.5.3 Initial storage and inflow sensitivity test

The initial storage at the beginning of the simulation is an input that needs to be specified for the model. The initial values can be prescribed from the observations, if available. However, the simulation of a hydrological/land surface model could start at any point in time when there is no observation to prescribe (e.g., some time in far past, a future scenario simulation, or a hypothetical scenario). Additionally, in a long-term simulation, the initial storage may result from a previous

model simulation, which may not be as close to observations as desired. The aim of the experiment is to examine and show to what extent the initial storage value affects the simulation performance.

To test the effect of initial storage used in the reservoir simulation performance, two experiments were conducted on three reservoirs with different scales of regulations: 1) Charvak (c=0.28), 2) Gardiner (c=1.46), and 3) High Aswan (c=2.84). In the first experiment, the initial storage was allowed to vary between ten percent of maximum storage ($0.1 \cdot S_{max}$) to maximum storage (S_{max}). In the second experiment, the initial storage range was narrowed to starting simulation month minimum and maximum historical observations. In both tests, 150 simulations were conducted by sampling the initial storage using uniform random sampling from the defined storage range.

Fig. 3.9 and Table 3.2 show the results of these initial storage perturbation experiments. For both experiments the simulations on the Charvak dam showed a similar range for NSE (flow) [0.79, 0.83] and NSE (storage) [0.61, 0.74]. Using one year as a spin-up period on Charvak dam simulations stabilized the initial storage effects, resulting in NSE (flow) of 0.82 and NSE (storage) of 0.74. The simulations on Gardiner dam in the first experiment showed a range of [0.35, 0.51] for NSE (flow) and [-0.43, 0.88] range for NSE (storage), while in the second experiment the ranges were narrowed to [0.44, 0.49] for NSE (flow) and [0.87, 0.88] for NSE (storage). For a one year spin-up period on the Gardiner dam this simulation converged the NSE (flow) range to [0.49, 0.51] and the NSE (storage) range to [0.76, 0.87] in the first experiment and to 0.49 NSE (flow) and 0.87 NSE (storage) for the second experiment. On the other hand, the simulation on the High Aswan dam showed a range of [-0.28, 0.85] for NSE (flow) and [0.38, 0.91] for NSE (storage) for the first experiment and [0.52, 0.85] for NSE (flow) and [0.42, 0.91] for NSE (storage) for the second experiment. Excluding a one year spin-up period from the metric calculation on the High Aswan dam simulation narrowed the NSE (flow) range to [0.62, 0.85] and the NSE (storage) range to [0.58, 0.91] for both experiments. Overall, as expected, the experiments suggest that the effect of initial storage on reservoir simulation performance depends on the regulation scale. Starting from observed storage values and using a one-year warm-up period allows stabilization of the

initial storage effect for low and medium regulated reservoirs. However, for highly regulated reservoirs, as in the case of High Aswan, longer spin-up periods are needed to stabilize the simulations. For example, a five-year spin-up period was required to fully stabilize the performance for the High Aswan dam simulations.

The existence of inflow bias is inevitable in any hydrological modelling practice. To understand the behaviour of the DZTR model under biased inflow conditions, we conducted a sensitivity experiment on the Charvak, Gardiner and High Aswan reservoirs. To do so, the DZTR model performance was tested using five simulations in which the entire inflow time series was changed by -50%, -25%, 0%, +25%, and +50%. The sensitivity of simulations to bias in inflow was evaluated using the NSE (flow) and NSE (storage) performance metrics.

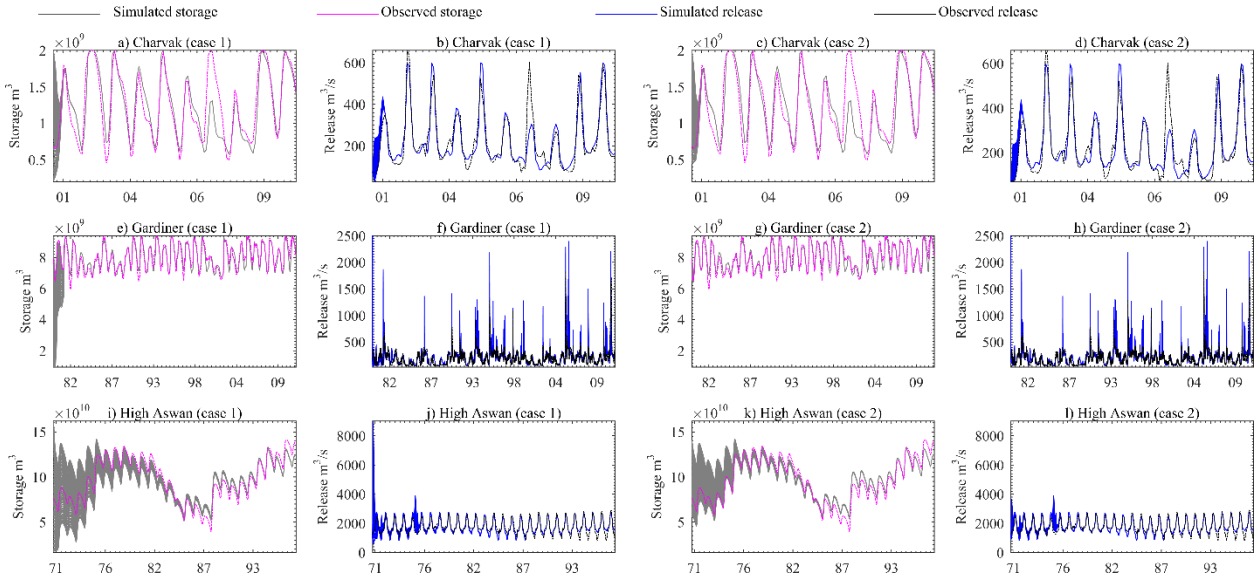


Figure 3. 9 Reservoir initial storage effect on storage and release simulation: (a) Charvak storage case 1, (b) Charvak release case 1, (c) Charvak storage case 2, (d) Charvak release case 2, (e) Gardiner storage case 1, (f) Gardiner release case 1, g) Gardiner storage case 2, (h) Gardiner release case 2, (i) High Aswan storage case 1, (j) High Aswan release case 1, (k) High Aswan storage case 2, (l) High Aswan release case 2

Fig. 3.10 and Table 3.3 show the results of the inflow bias test and that the reservoir simulation performance significantly changes as a result of this bias. Reducing the inflow by 50% considerably reduced the reservoir storage and release and led to negative values of NSE (flow) and NSE (storage) for all reservoirs. For such a large negative inflow bias, the reservoir operation tries to recover the storage to the target (observed) level by releasing as low as possible. Conversely, the positive inflow bias increased simulated storage and releases for all reservoirs, which led to negative performance metrics for all reservoirs except on Gardiner NSE (storage). As shown in **Fig. 3.10**, with large positive inflow bias, storage quickly moves towards flood and maximum storage targets resulting in insufficient storage left to attenuate flood peaks and the operation model starts discharging large releases through the spillway to maintain the storage at the maximum storage target. Inflow bias of -25% and +25% showed similar behaviour as -50% and +50% bias for all reservoirs, but the simulation performance metrics during -25% and +25% provide significant positive NSE values for the Charvak and Gardiner dams except for the Gardiner NSE (flow) for +25% which resulted a negative NSE value. However, on the highly regulated High Aswan dam, the ±25% inflow bias significantly reduced the performance to negative values.

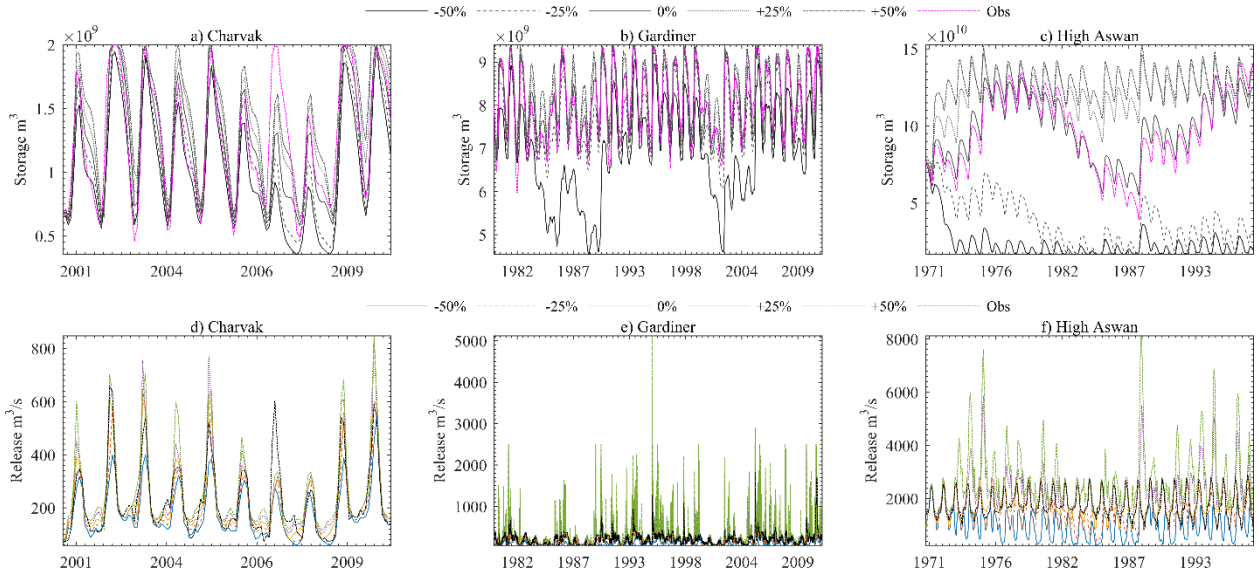


Figure 3. 10 Inflow bias sensitivity test on storage and release simulation: (a) Charvak storage, (b) Gardiner storage, (c) High Aswan storage, (d) Charvak release, (e) Gardiner release, (f) High Aswan release. The *x* axes show time and the *y* axes show release and storage values.

Table 3. 2 Reservoir initial storage effect on storage and release simulation

		Case 1 S_0**= [0.1**S_{max} S_{max}**]**		**Case 2** S_0**=[min(obs) max(obs)] obs= observed for all Jan 1st**	
		NSE(Storage)	NSE(Flow)	NSE(Storage)	NSE(Flow)
Charvak	No spin-up	[0.61 0.74]	[0.79 0.83]	[0.61 0.74]	[0.79 0.83]
	1yr spin-up	[0.74 0.74]	[0.82 0.82]	[0.74 0.74]	[0.82 0.82]
Gardiner	No spin-up	[-0.43 0.88]	[0.35 0.51]	[0.87 0.88]	[0.44 0.49]
	1yr spin-up	[0.76 0.87]	[0.49 0.51]	[0.87 0.87]	[0.49 0.49]
High Aswan	No spin-up	[0.38 0.91]	[-0.28 0.85]	[0.42 0.91]	[0.52 0.85]
	1yr spin-up	[0.58 0.91]	[0.62 0.85]	[0.58 0.91]	[0.62 0.85]

Table 3. 3 Inflow bias sensitivity test on storage and release simulation

		-50%	**-25%**	**0%**	**25%**	**50%**
Charvak	NSE(Storage)	-1.95	0.25	0.74	0.52	-0.21
	NSE(Flow)	-0.06	0.54	0.82	0.57	-0.07
Gardiner	NSE(Storage)	-2.00	0.74	0.88	0.79	0.66
	NSE(Flow)	-0.21	0.47	0.49	-0.43	-2.02
High Aswan	NSE(Storage)	-9.37	-5.96	0.90	-0.60	-1.45
	NSE(Flow)	-3.90	-0.34	0.80	-2.29	-8.70

3.5.4 Parameter calibration and validation of the DTZR model

We tried to improve upon the generalized parameterization by calibrating the DZTR parameters via bi-objective optimization for two objective functions, Nash-Sutcliffe on reservoir storage (NSE (storage)) and Nash-Sutcliffe on reservoir release (NSE (flow)). This is an important step when the data and computational resources for optimization are available, to enhance reservoir simulation and consequently hydrological modelling of the region of interest. **Fig. 3.11** shows the multi-criteria reservoir calibration (yellow circles) and validation (red circles) Pareto solutions for all reservoirs. The Pareto solutions show strong tradeoffs between fitting observed reservoir storage versus downstream release, which also reflects the fact that the problem is multi-objective by nature and it is required to consider both storage and release, instead of fitting one at the cost of degrading the other. The generalized parameterization solution for the calibration (yellow

square with blue border) and validation periods (red square with blue border) is also added in **Fig. 3.11** for each reservoir to show the improvement gained through parameter calibration. Relative to the generalized solution for the calibration period, reservoir parameter calibration improved both NSE (flow) and NSE (storage) for all reservoirs with a median improvement of 0.11 and 0.21, respectively. The NSE (flow) improvement ranged from 0.017 to 0.575, and NSE (storage) improvement ranged from 0.02 to 0.66. The parameter calibration has shown significant improvement on reservoirs that have lower performance with generalized parameterization. The best examples of this case are Fort Randall, Int. Amistad, Trinity, Int. Falcon, and E.B. Campbell, as shown in **Fig. 3.11**. Small improvements in performance have also been observed on reservoirs that have greater performance with generalized parameterization such as American Falls, Andijan, Nurek, High Aswan, Waterton, and Charvak. The validation of calibrated solutions improved the NSE (flow) and NSE (storage) for 56% of the reservoirs with a median improvement of 0.035 and 0.092, respectively. The NSE (flow) improvement in the validation period ranged from 0.001 to 0.335, and NSE (Storage) improvement ranged from 0.004 to 1.02. During validation, the remaining reservoirs (44% of them) resulted in NSE (flow) and NSE (storage) reductions with a median reduction of 0.032 and 0.089, respectively. The reductions of NSE (flow) ranged from 0.001 to 0.073, and those of NSE (storage) ranged from 0.001 to 0.257.

Overall, considerable improvement was achieved for both calibration and validation periods for several reservoirs such as the Dickson, Gardiner, Ghost, Int. Amistad, Int. Falcon, Kayrakkum, Sirikit, Yellowtail, and Glenmore. However, as shown in **Fig. 3.11**, the improvements of DZTR model performance during calibration do not usually guarantee performance improvement in validation. This is because, as for any other types of model as well, the properties of the calibration and validation periods might differ significantly. In particular, the calibrated Pareto solution does not show the same trade-off or level of performance during validation when there is considerable change in inflow properties as a result of consecutive wet or dry years. Examples of this condition are shown for Glen Canyon (similarly Bhumibol, Fort Randall, and Fort Peck) where the calibration period had more wet and high inflow years than the validation

period. Such considerable changes of inflow, storage, and release results in performance degradation during the validation period. In general, a small change in inflow, storage, or release for the validation period can change the shape of the trade-off. However, the calibrated parameters in most cases were still capable of producing good performance during validation close to, or better than, that of the generalized parameterization for the same period.

To further test the role of the calibration period, we calibrated all reservoirs using the whole observational record. The result of this test is shown in **Fig. 3.12,** which demonstrates the strong role of the calibration period. All reservoirs showed trade-off between storage and release fitting. The solution resulted in a consistent Pareto pattern similar to the split-sample calibration results. The median NSE (flow) and NSE (storage) improvement when using the whole observational record for calibration are approximately 0.1 and 0.12 respectively, while the maximum improvement reached 0.45 and 0.55 for some reservoirs. High improvements on storage and flow simulations in the case of whole-period-calibration are mostly observed on reservoirs that have considerable shift of observed storage and flow across the period of observation period. **Fig. 3.13** shows some example reservoirs that had considerable improvements such as Bhumibol, Canyon Ferry, Int. Amistad, Int. Falcon, Navajo, and Trinity dams, compared to generalized parameters (**Fig. 3.6**). Similarly, for the remaining reservoirs, calibrating the whole period showed (**Fig. 3.13**) better agreement of daily and monthly simulations with the observations, even for years with extreme deviations that are most likely associated with extreme dry and wet conditions. Additionally, the long-term average simulations (**Fig. 3.14**) showed that calibrating using the whole period reduced the deviation between simulations and observations, and in most cases the Pareto simulation range encompasses the observation. Overall, the calibration period test indicates the benefit of using long-term observation for parametrization (even for generalized parameterization) to allow the parametrization to represent behaviour in extreme periods. Thus, we recommend using as much data as available to parameterize the model for a specific reservoir so that all information on reservoir operation will be accounted for.

The DZTR scheme introduces more parameters to the host land surface model. However, its parameters are external to those of land surface model and are determined a priori using storage and release data. The decision of the time scale to use for specifying the parameters is left to the modeller. The user has the ability to investigate the seasonal patterns in the storage and release data and decide whether a monthly or a coarser time scale (e.g., quarterly) would be sufficient. In fact, the configuration of DZTR is also flexible to use any user-specified zoning that are available from observation, reservoir information or zoning values specified in other studies such as *Zhao et al.* (2016).

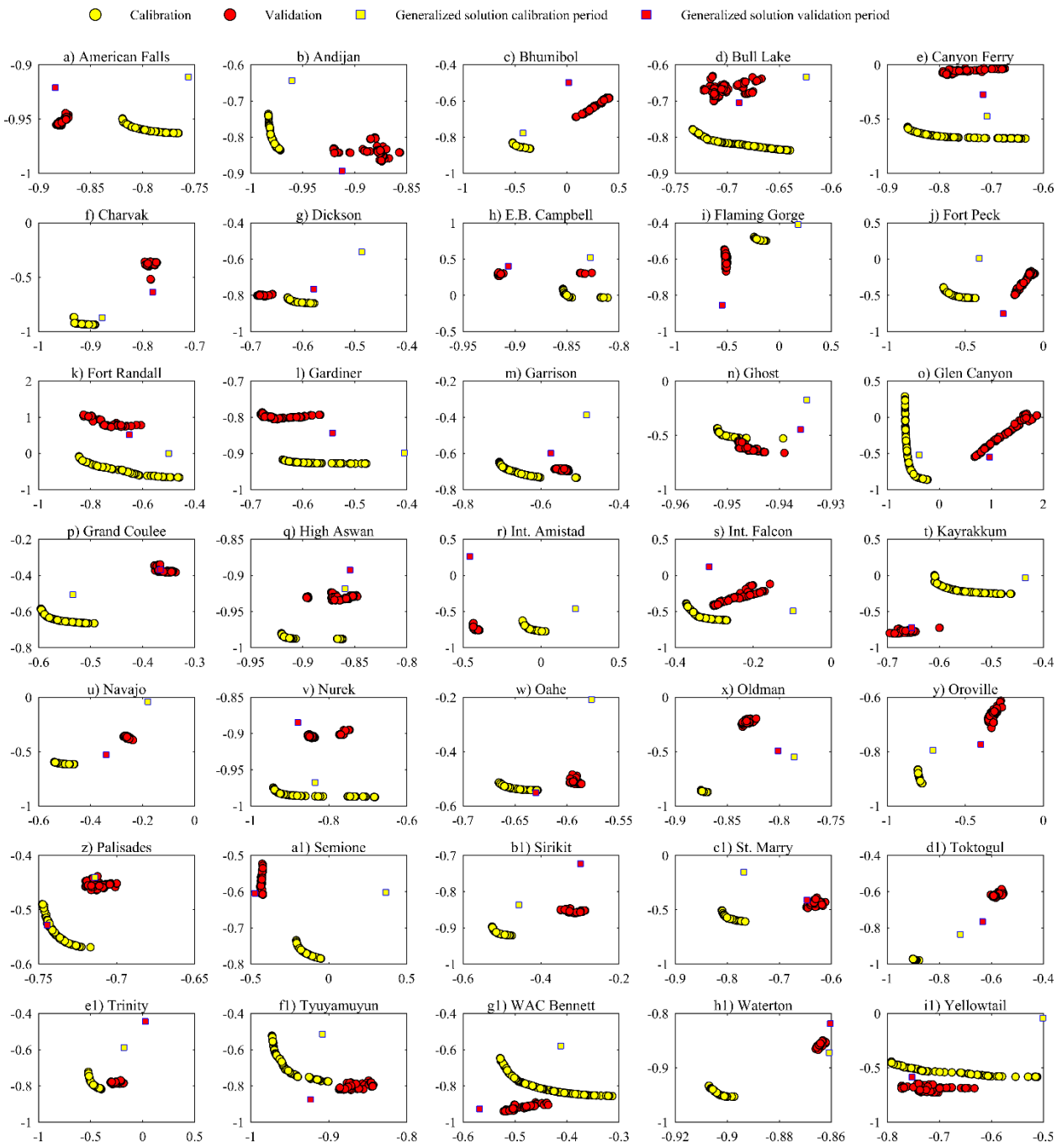


Figure 3. 11 Reservoir release parameter multi-objective calibration result, the *x* axes show NSE (flow) multiplied by -1 and the *y* axes show NSE (storage) multiplied by -1.

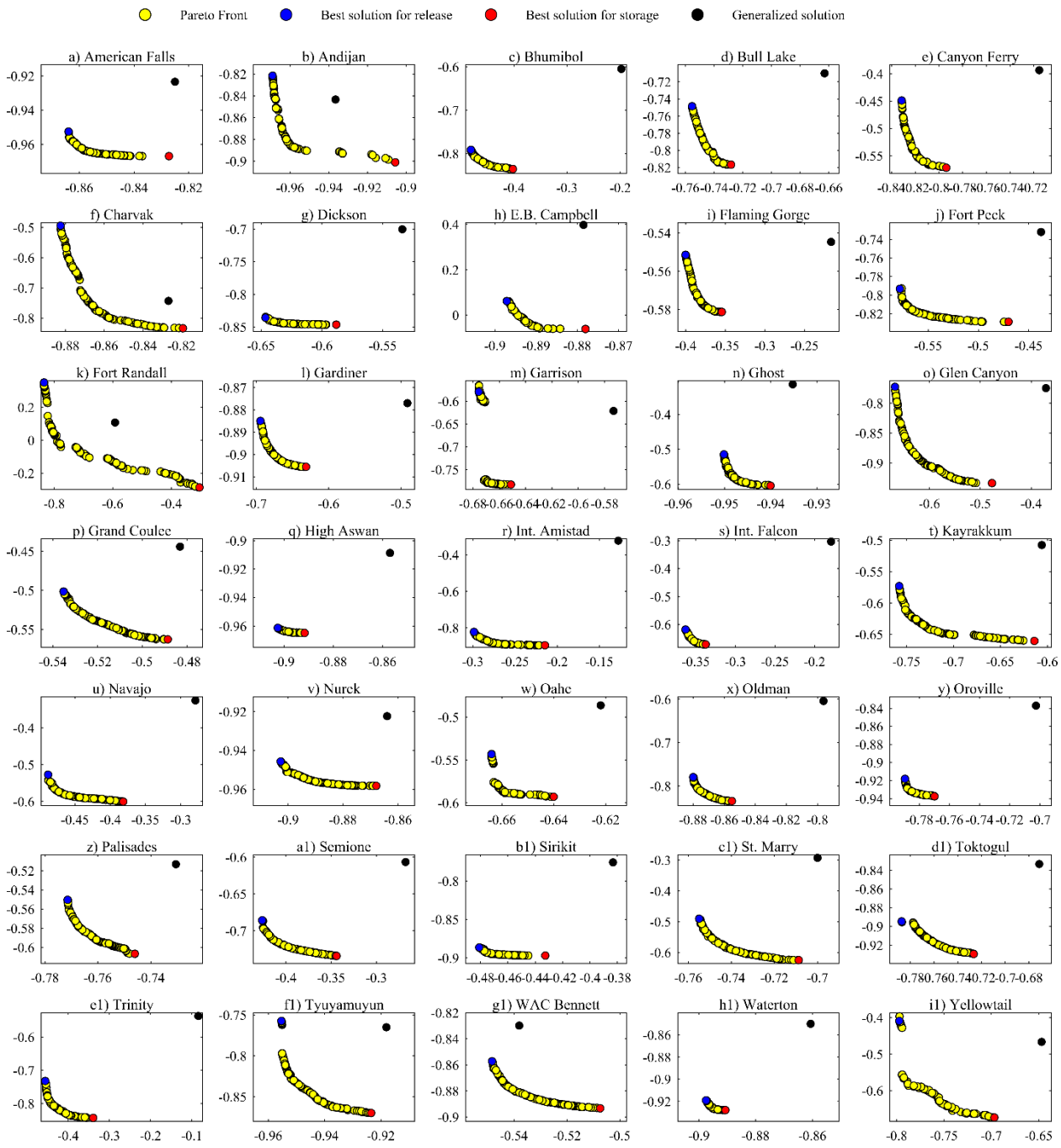


Figure 3. 12 Reservoir release parameter multi-objective calibration using all available data for each reservoirs, *x* axes show NSE (flow) multiplied by -1 and the *y* axes show NSE (storage) multiplied by -1.

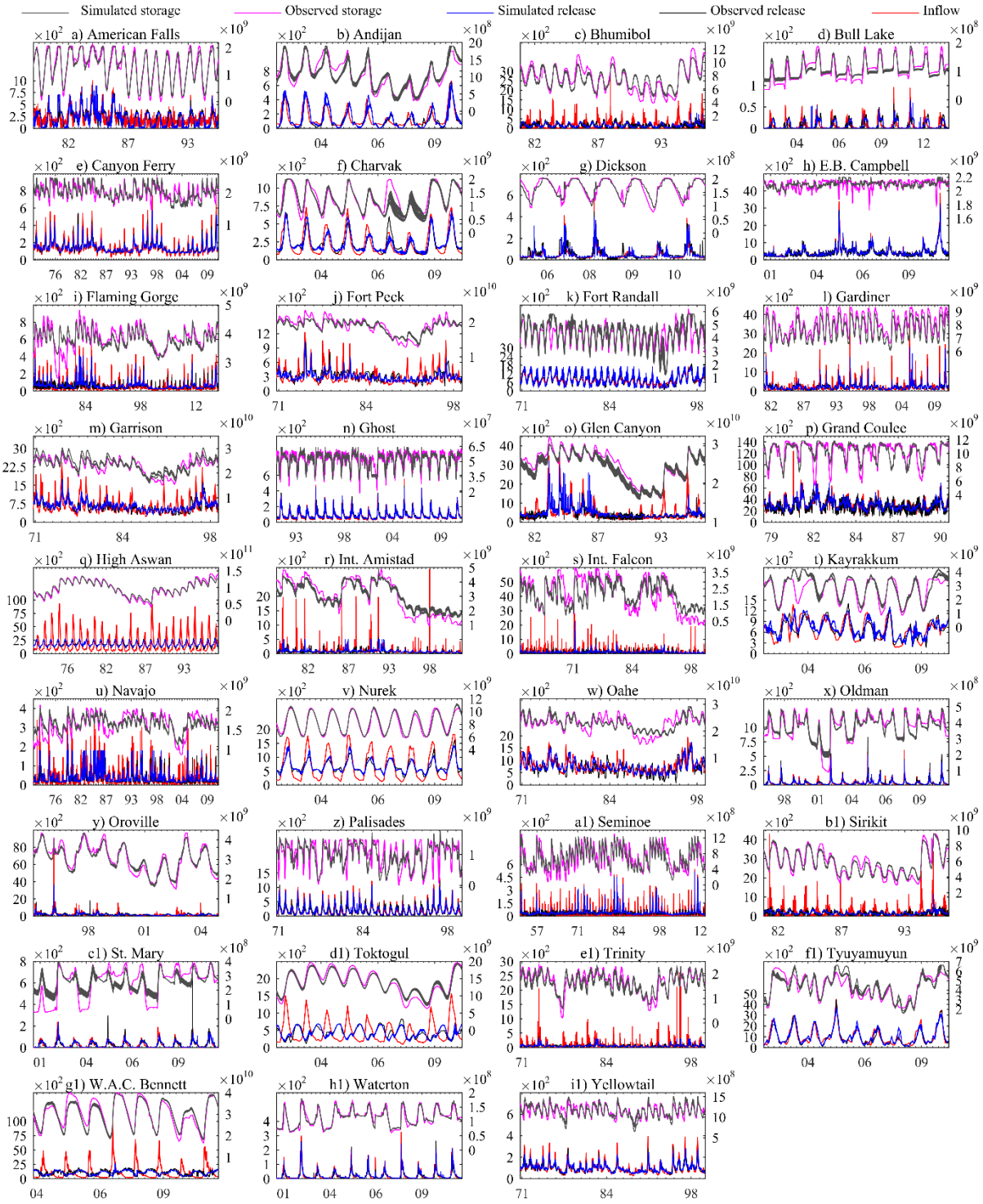

Simulated storage
Observed storage
Simulated release
Observed release
Inflow
a) American Falls
b) Andijan
c) Bhumibol
d) Bull Lake
e) Canyon Ferry
f) Charvak
g) Dickson
h) E.B. Campbell
i) Flaming Gorge
j) Fort Peck
k) Fort Randall
l) Gardiner
m) Garrison
n) Ghost
o) Glen Canyon
p) Grand Coulee
q) High Aswan
r) Int. Amistad
s) Int. Falcon
t) Kayrakkum
u) Navajo
v) Nurek
w) Oahe
x) Oldman
y) Oroville
z) Palisades
a1) Seminoe
b1) Sirikit
c1) St. Mary
d1) Toktogul
e1) Trinity
f1) Tyuyamuyun
g1) W.A.C. Bennett
h1) Waterton
i1) Yellowtail

Figure 3. 13 Daily and monthly reservoir simulations using DZTR model with a generalized parametrization, the *x* axes show month or year, the primary *y* axes show release (m^3 s^{-1}) and the secondary *y* axes show storage (m^3).

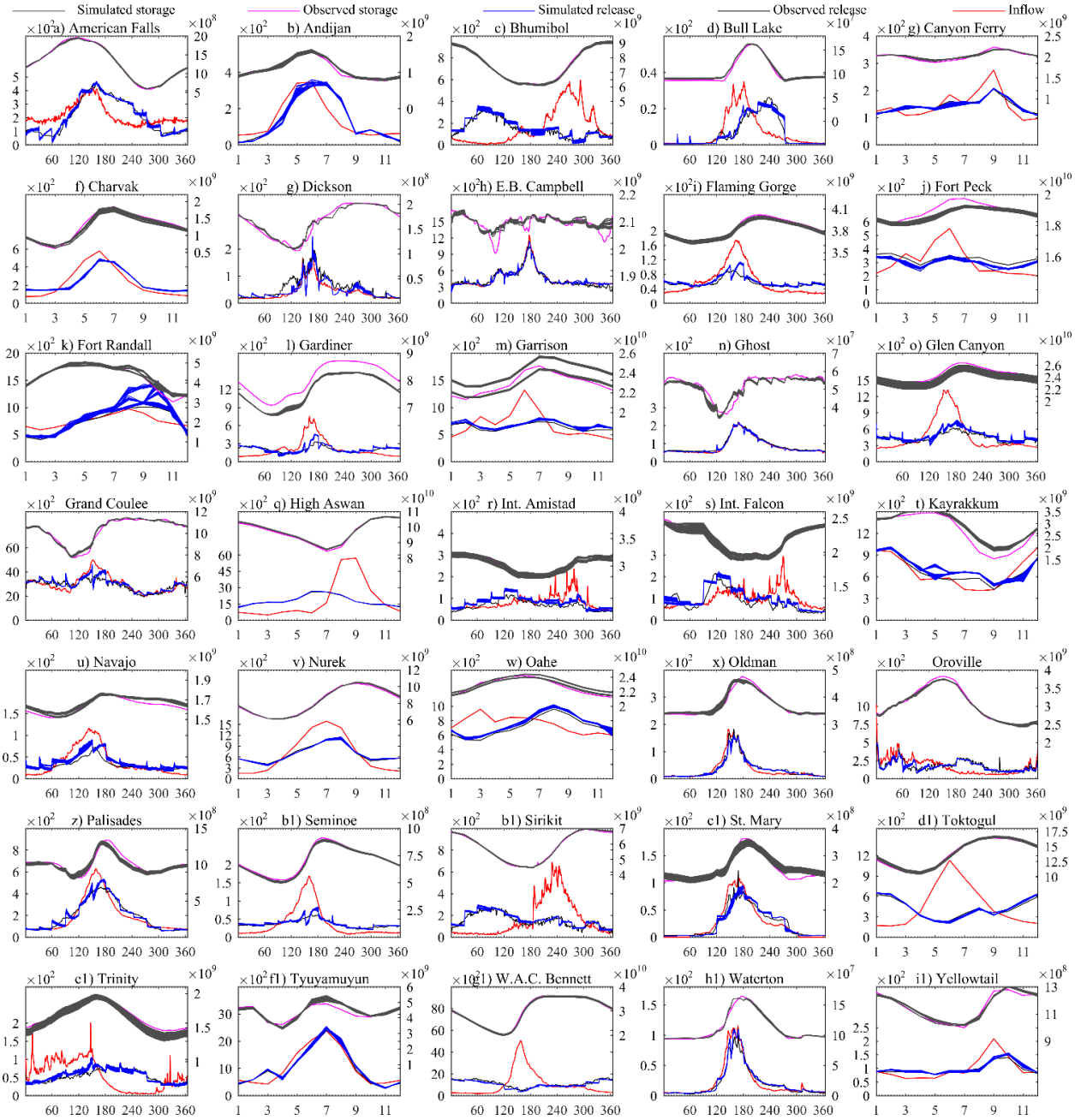


Figure 3. 14 Long-term average daily or monthly reservoir simulations with generalized parametrization, the *x* axes show days (1-365) or months, (1-12) the primary *y* axes show release (m^3 s^{-1}) and the secondary *y* axes show storage (m^3).

3.5.5 DZTR model test within the MESH model

Finally, the generalized parametrization of the DZTR model was integrated into the MESH model and tested to simulate six reservoirs in the Saskatchewan River Basin (Gardiner, St. Mary, Waterton, Oldman, Ghost and Dickson dams) and one reservoir (W.A.C. Bennet dam) in the Mackenzie River Basin, both in Western Canada. The reservoir simulation was run using MESH-modelled inflows at a half-hourly time step, the usual MESH modelling time step, and the performance metrics were calculated at a daily time step. The MESH-modelled inflows are considered to represent the base-case scenario, and the inflow can be assumed as regulated or natural depending on whether there are dams upstream or not.

Fig. 3.15 illustrates that the generalized DZTR model generally improves upon having no representation of the reservoirs in the model. This improvement is apparent in the NSE (flow) values which increase with the DZTR model. The exception is Dickson dam with a small reduction in NSE. The importance of integration of the DZTR model was predominant for the Gardiner and W.A.C. Bennett dams, which are highly regulated (c>0.5) when compared to the other reservoirs tested in MESH.

This general improvement of flow simulation when comparing a reservoir model to the no-reservoir assumption is, of course, not surprising. What is important to note, however, is that the improvement in NSE can be substantial without calibration of the DZTR parameters. This is important for many LSM applications where calibration is generally not performed. *Hanasaki et al.* (2006) illustrate that their method is superior to the natural lake (or unregulated reservoir) method applied in many CMs and H-LSMs, and this paper shows that the DZTR model improves upon the results of *Hanasaki et al.* (2006). Therefore, it is natural to assume that the DZTR model would also be an improvement in uncalibrated H-LSM applications.

However, calibration is very common in CM or H-LSM applications in which the DZTR model would likely be employed. A full comparison of calibrated results between a no-reservoir case, natural lake (or unregulated reservoir), and the DZTR model (and the other reservoir models) is beyond the scope of this paper. Again, given the improvements shown with the uncalibrated DZTR model when compared with other uncalibrated models, and the general improvements shown here when calibrating the DZTR model, it is assumed that calibrating the DZTR model within a CM or H-LSM would improve upon calibrating an unregulated reservoir model, or the other reservoir models compared in this paper. This general improvement of flow simulation when comparing a reservoir model to the no-reservoir assumption is, of course, not surprising. What is important to note, however, is that the improvement in NSE can be dramatic without calibration of the DZTR parameters. This is important for many LSM applications where calibration is generally not performed. *Hanasaki et al.* (2006) illustrated that their method is superior to the natural lake (or unregulated reservoir) method applied in many CMs and H-LSMs, and this paper shows that the DZTR model improves upon the results of *Hanasaki et al.* (2006). Therefore, it is natural to assume that the DZTR model would also be an improvement in uncalibrated H-LSM applications.

The storage simulation showed low NSE (storage) value for the St. Mary and Waterton dams and negative NSE (storage) for Oldman and Ghost dams. However, the simulation showed a reasonable representation of storage variability, but with considerable underestimation. This underestimation in storage in **Fig. 3.15** is attributable to the fact that the modelled inflow is underestimated. It is expected that calibration of the land-surface parameters in conjunction with the DZTR parameters in MESH would improve the modelled inflows and resulting modelled reservoir storage.

It is worth mentioning again that H-LSMs, such as MESH, can also be used for the original purpose of LSMs, which is to represent fluxes from the land-surface to the atmosphere. If the approach improves modelled flows where reservoirs operate, it could result in a better parameterization of the LSM, which should in-turn improve land-surface fluxes and feedbacks to the atmosphere.

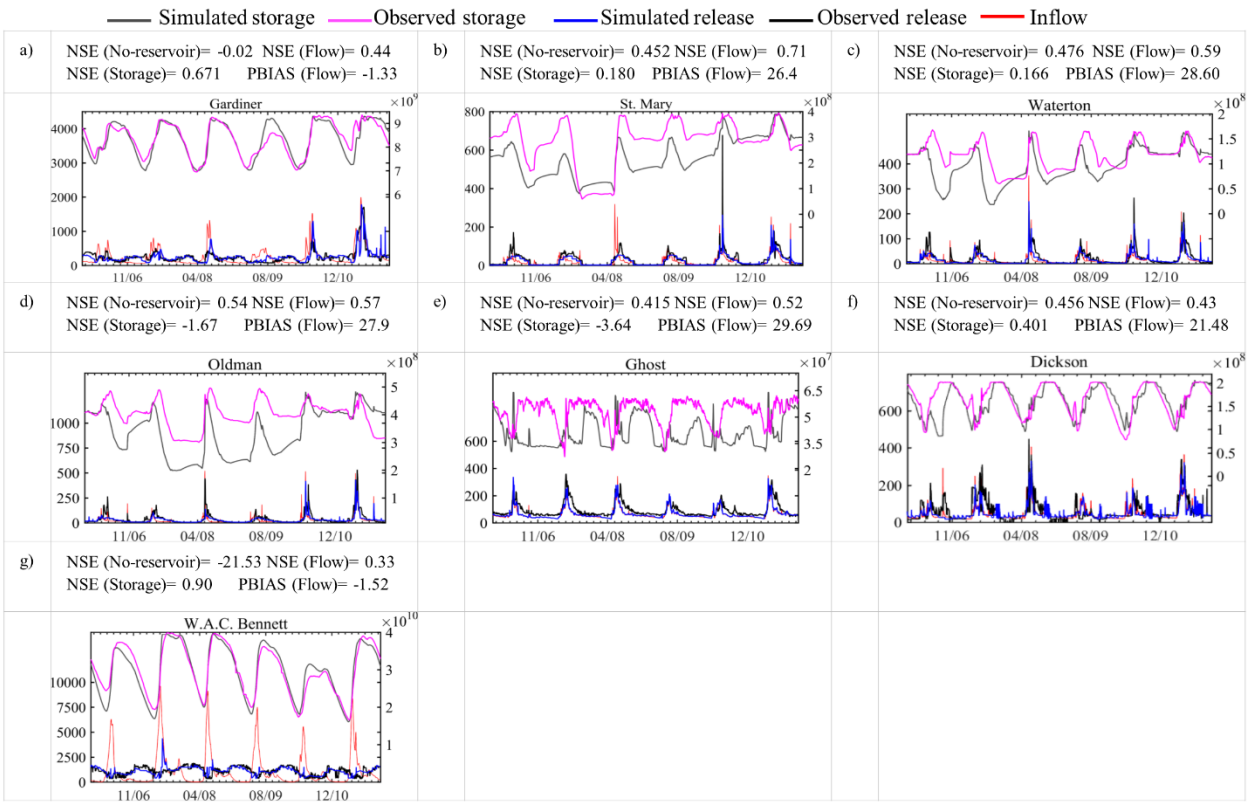


Figure 3. 15 Reservoir simulation results within MESH model run for selected reservoirs. The *x* axes show time, the primary *y* axes show release (m^3 s^{-1}) and the secondary *y* axes show storage (m^3).

3.5.6 Uncertainties in reservoir operations and DZTR parametrization

Reservoir operation involves considerable uncertainties that may arise due to several factors. One major source of uncertainty in reservoir operations is future inflows (long-term and short-term inflow forecast). The forecast contains errors rooted in the forecast method, the driving climate forecast, snowpack measurements, timing of snowmelt and the statistical (stationarity) assumptions to generate inflows based on historical inflows. The inflow forecast uncertainty is more significant during flood seasons because it involves subjective decisions of operators to avoid the risk of dam overtopping and downstream flooding. Other sources of uncertainty in reservoir

operations include changes in demand over time because of increases in demand for irrigation, power, water supply, etc. The purpose of the reservoir can also change from its initial intended purpose (e.g., adding a hydropower station to an irrigation dam). These changes are only implicitly captured by the DZTR scheme as implied in the storage and release time series used for parameterizing it for a specific reservoir.

Given the above uncertainties, even the actual reservoir operation may deviate from the designed reservoir operation rule curve. Some of the decisions of reservoir operators are spontaneous, ad-hoc, and depend on experiences that are not usually documented. Thus, there are difficulties to accurately represent the historical operation or to establish accurate relationships between reservoir storage, inflows, and releases. These relationships typically contain considerable noise e.g., different release values for the same storage level during the same season. As a result, these uncertainties influence considerably the parameterization of the model derived to represent the reservoir operation based on historical observations of each reservoir. This is particularly true for the algorithm presented because of two main factors. Firstly, the presented reservoir algorithm assumes that the relationship between reservoir storage and releases follow piecewise linear functions. There is a chance that other functional forms represent such relationships better for some reservoirs. Secondly, in the case of the generalized parameterization, the bending points in the piece-wise linear functions (zone classification points) are estimated based on fixed probabilities of exceedance extracted from historical data for all reservoirs. A different dataset (of reservoirs and/or time periods) could result in different quantiles. The assumption of having similar bending points of the piecewise linear functions for all reservoirs cannot provide optimal zones for each reservoir. However, we showed that the generalized parameterization performs better compared to other widely used algorithms.

Optimizing storage and release parameters allows to overcome the limitation of generalized bending points of the piecewise linear function by adjusting the bending points so that the best fit can be identified. However, optimization usually does not provide a perfect storage release relationship (i.e., in general, the trade-off between objectives never converges to a single point),

because the perfect representation only occurs in the case of a perfect reservoir model and perfect data. The proposed model, like many other types of models is not an exception because of the uncertainties highlighted in the previous point. Thus, the trade-off between storage and release objectives can be viewed as a measure of the limitation of the reservoir algorithm (piece-wise linear functions, fixed number of zones, etc.) and observation errors. To examine the level of uncertainty of the trade-off, it is important to look at the shape and range of the trade-off on each objective function axis.

Apart from a few reservoirs, the range of Pareto solutions for each objective function is generally narrow with good NSE values (**Fig. 3.11** and **Fig. 3.12**). In such cases, the associated uncertainties are less and the trade-off between improving simulated releases and improving simulated storage is minimal. Conversely, in some cases, an extended spread of the trade-off along one of the axes (objective function) was observed, indicating a higher uncertainty of the algorithm for the process that the axis represents, i.e., reservoir storage or release. This requires further investigations of the datasets and parameterization for those reservoirs and their history of operations. Shifts in operational management of reservoirs do occur and these may obscure the parameterization. These may be detected by careful examination of the available records as well as metadata records of the reservoir history if accessible. The level of noise when determining the parameters could be an indicator of changes in operation.

3.5.7 Implementation strategies to overcome data limitation

The data requirement is the main limitation of the DZTR model for application at continental and global scales. One approach to overcome data limitations is to integrate our proposed method in land surface/catchment models along with other reservoir operation methods (e.g., *Hanasaki et al.*, 2006). Then, within the land surface/catchment models, identifier flags can be used to indicate which method applies to which reservoirs. The DZTR approach can only be activated for reservoirs

with data support, while the remaining reservoirs can use other approaches as dictated by data availability. We have been following such an implementation within the MESH model.

As shown in our results, reservoir regulation has a large impact on downstream flows if the reservoir is highly regulated and/or is of multi-year type (c>0.5). Thus, more emphasis can be placed on those reservoirs with c>0.5. At the moment, such methods will be more effective at the regional than global scale (for example for the Saskatchewan River Basin in our case), because modellers at regional scales have better access to inflow-storage-outflow data and have better understanding of the system to acquire the necessary reservoir data. In a land surface hydrologic model, important reservoirs are those causing large changes to the downstream flows and those tend to be the larger ones with generally better data availability.

Data on reservoir storage, inflows and releases exist for most reservoirs but sometimes they are not made publically available. Storage data can be obtained from water level data, which are generally available for major reservoirs and can be converted to storage. Release data can be deduced from the nearest downstream station. In addition, new initiatives are needed to gather and archive such reservoir datasets and move beyond information on reservoir characteristics that is currently available in databases (e.g., GRanD database: *Lehner et al.*, 2011). One of our recommendation is that the target release and storage data be archived for public use at least for highly regulated and multi-year type dams (c>0.5).

The possibility of estimating storage and release data from different satellite data products is promising. Such new data sources will potentially improve the use of methods like the presented reservoir operation (optimized or generalized). More recently, *Busker et al.* (2019) showed estimation of volume for 130 reservoirs using a surface water dataset and satellite altimetry, this is an encouraging approach to reduce the data limitation.

3.6 Summary and conclusions

Human interventions in hydrologic systems through dams and reservoirs significantly change the flow regime of many rivers. In this paper, we presented an improved reservoir operation model, called Dynamically Zoned Target Release (DZTR) model that can be integrated into any large-scale hydrological model; here we integrated it into the MESH hydrology-land surface model. The DZTR model is based on parametric piecewise linear functions that approximate reservoir release rules used by reservoir operators. We proposed two strategies to identify the parameters of this model: one based on the distributions of historical storage and release to generate the so-called "generalized parameters" and the other one based on direct calibration to observed storage and release time series via multi-objective optimization. We first tested the DZTR model individually across a number of reservoirs around the globe, and then tested its performance when plugged into the MESH model for a subset of those reservoirs. Our conclusions can be summarized as:

- The DZTR reservoir operation model performed well in reproducing observed storage and release time series in (almost) all reservoirs tested and outperformed the existing reservoir models proposed by *Hanasaki et al.* (2006) and *Wisser et al.* (2010). The model was capable of capturing inter- and intra-annual variability of both reservoir storage and release.
- As expected, calibration significantly improved the performance of the DZTR model compared with the performance of the "generalized parameters". There often exists, however, a significant tradeoff between fitting reservoir storage versus release, signifying the importance of accounting for both storage and release in a multi-objective fashion.
- The integration of the DZTR reservoir model into the MESH hydrology-land surface modelling system was straightforward and improved the overall model performance compared with the traditional methods of accounting for reservoirs in H-LSMs. This integration can be viewed as a successful example for improving the representation of reservoir operation in CMs, LSMs and GWSMs.

Future research work may include (1) examining the applicability of the DZTR model for regions with severely-limited data by examining the utility of other data sources such as those derived from satellite-based observations (*Savtchenko et al.*, 2004; *Garambois and Monnier*, 2015; *Gao et al.*, 2012) and using an area-volume relationship approximated by regular geometric shapes (e.g., *Yigzaw et al.*, 2018); and (2) examining a direct one- and/or two-way coupling of WMMs with CMs and LSMs towards developing a seamless coupled framework for the simulation of natural-engineered watershed systems.

CHAPTER 4

HYDROLOGIC-LAND SURFACE MODELLING OF A COMPLEX SYSTEM UNDER PRECIPITATION UNCERTAINTY: A CASE STUDY OF THE SASKATCHEWAN RIVER BASIN, CANADA

This chapter is based on the following published article under discussion.

Yassin, F., Razavi, S., Wong, J. S., Pietroniro, A. and Wheater, H.: Hydrologic-Land Surface Modelling of a Complex System under Precipitation Uncertainty: A Case Study of the Saskatchewan River Basin, Canada, *Hydrol. Earth Syst. Sci. Discuss.*, 1–40, doi:10.5194/hess-2019-207, 2019b. https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-207/

Author contributions: FY developed the model setup and conducted the conceptualization and simulations. FY analyzed the results and primarily wrote the manuscript. All co-authors critically reviewed the manuscript.

Synopsis

Hydrologic-Land Surface Models (H-LSMs) have been progressively developed to a stage where they represent the dominant hydrological processes for a variety of hydrological regimes and include a range of water management practices, and are increasingly used to simulate water storages and fluxes of large basins under changing environmental conditions across the globe. However, efforts for comprehensive evaluation of the utility of H-LSMs in large, regulated watersheds have been limited. In this study, we evaluated the capability of a Canadian H-LSM, called MESH, in the highly regulated Saskatchewan River Basin (SaskRB), Canada, under the constraint of precipitation uncertainty. The SaskRB is a complex system characterized by hydrologically-distinct regions that include the Rocky Mountains, Boreal Forest, and the Prairies. This basin is highly vulnerable to potential climate change and extreme events. A comprehensive analysis of the MESH model performance was carried out in two steps. First, the reliability of multiple precipitation products was evaluated against climate station observations and based on their performance in simulating streamflow across the basin when forcing the MESH model with a default parameterization. Second, a state-of-the-art multi-criteria calibration approach was applied, using various observational information including streamflow, storage and fluxes for calibration and validation. The first analysis shows that the quality of precipitation products had a direct and immediate impact on simulation performance for the basin headwaters but effects were dampened when moving downstream. In particular, the Canadian Precipitation Analysis (CaPA) performed the best among the precipitation products in capturing timings and minimizing the magnitude of errors against observations, despite a general underestimation of precipitation amounts. The subsequent analyses show that the MESH model was able to capture observed responses of multiple fluxes and storage across the basin using a global multi-station calibration method. Despite poorer performance in some basins, the global parameterization generally achieved better model performance than a default model parameterization. Validation using storage anomaly and evapotranspiration generally showed strong correlation with observations, but revealed potential deficiencies in the simulation of storage anomaly over open water areas.

4.1 Introduction

During the past few decades, Land Surface Models (LSMs) have expanded in scope and complexity. As they have become more sophisticated, they have increasingly integrated dominant hydrological processes, such as horizontal hydrological fluxes, subsurface lateral water movement, and river flow routing, all of which are well recognized in the Hydrological Models (HMs) (*Archfield et al.,* 2015; *Davison et al.,* 2016). More recently, several LSMs have included irrigation and water management modules (*Voisin et al.* 2013a, 2013b; *Haddeland et al.,* 2006; *Pokhrel et al.,* 2017), which are well established in Global Hydrological Models (GHMs) (*Döll et al.,* 2003; *Wada et al.,* 2017; *Archfield et al.,* 2015). The integration of these various processes has enabled LSMs to be used in support of a wide range of hydrological applications, in which they are referred to as Hydrologic-Land Surface Models (H-LSMs) (*Pietroniro et al.,* 2007). Although H-LSMs have made steady advances in representing hydrologic processes and incorporating human impacts on the terrestrial water cycle, the investigation of input data uncertainty and parameter estimation through calibration for large-scale basins has been limited and is not common practice with H-LSM models compared to their extensive use by the catchment hydrological modelling community. However, quantifying the sources of uncertainty and their magnitude associated with inputs and model simulations is essential to ensure the effective use of the H-LSMs and increase the reliability of model predictions, especially for large-scale basins that are increasingly influenced by human activities. This paper addresses this important issue by evaluating the input and parameter uncertainty of an H-LSM with the use of multiple data sources and advanced calibration methods, using the Saskatchewan River Basin in western Canada, as a case study.

H-LSMs require large numbers of parameters to describe vegetation, soil and snow processes. The parameterizations of H-LSMs are often predefined by referring to look-up tables (*Mendoza et al.,* 2015) and are rarely calibrated, particularly for large-scale basins and global scale modelling (*Gupta et al.,* 1999; *Davison et al.,* 2016). The limited application of parameter calibration is largely due to the difficulties of managing the large number of H-LSM parameters and the high computational requirements. Such difficulties escalate when a model is applied to a

large-scale basin with complex surface heterogeneities and complex hydrologic and water management features. Possible ways to deal with the challenges of computational burden during calibrating H-LSMs are to reduce the number of parameters (*Houser et al.,* 2001; *Nasonova et al.,* 2009), use state-of-the-art computationally-efficient calibration methods (*Tolson and Shoemaker* 2007), and pre-emption strategies (*Razavi et al.,* 2010), and surrogate modelling (*Razavi et al.,* 2012).

Recently, attention has been given to advance H-LSM parameter estimation through model calibration, but model calibration and fidelity assessment are commonly based on streamflow observations only and the models have been mainly applied to smaller basins (*Nasonova et al.,* 2009). Calibration only to streamflow can mask major model deficiencies, as different components, such as model structure, model parameters and forcing data can compensate for the shortcomings, so that improvements to streamflow simulation calibration may be at the cost of degrading other model outputs. A possible approach to address this issue is to include other sources of information available in addition to streamflow during calibration and validation (*Crow et al.,* 2003; *Yassin et al.,* 2017; *Lo et al.,* 2010). For this purpose, remote sensing and local in situ observations can be used, such as the GRACE total storage anomaly (*Tapley et al.,* 2004), MODIS-based evapotranspiration and land surface temperature (*Mu et al.,* 2007; *Zhang et al.,* 2010), and evapotranspiration estimates based on flux tower latent heat flux observations (*Barr et al.,* 2012).

Additional issues arise due to input data uncertainty, which again becomes more important for larger scale applications. Most H-LSMs use energy-based approaches that require high spatio-temporal resolution climate forcing data. The spatio-temporal forcing data required increases the significance of input uncertainties and their impact on the quality of H-LSM outputs. Although observational climate station data are often regarded as the "truth", they have limited functionality (applicability) in driving H-LSMs for large basins. The density of meteorological station observations is too sparse and the temporal resolution too coarse to sufficiently represent the spatiotemporal variability of climate forcing variables (*Clark and Slater,* 2006). Additionally, climate station measurements are prone to precipitation under-catch error, particularly for solid

precipitation (*Mekis and Hogg,* 1999; *Adam and Lettenmaier,* 2003). The feasible option often chosen to drive H-LSMs for large-scale basins is to use gridded reanalysis climate forcing data derived from numerical weather prediction model output (*Sheffield et al.,* 2006; *Weedon et al.,* 2014; *Côté et al.,* 1998; *Mesinger et al.,* 2006) or use interpolated climate forcing based on climate station observations and radar data (*Mahfouf et al.,* 2007).

Although reanalysis and interpolated data provide better coverage of the space-time field of the meteorological conditions, they also contain significant errors that vary among products. The accuracy of each climate forcing product across different regions can have substantial impacts on the simulation of streamflow. For instance, *Eum et al.* (2014) showed that the runoff from three high-resolution precipitation datasets was significantly different during the snowmelt period over high elevation alpine areas in the Athabasca River basin. Hence, the evaluation of the quality of available reanalysis climate forcing datasets over a given region is an important prerequisite not only to select the best performing forcing data but also to reduce error propagation in both H-LSM outputs and parameter estimation (*Eum et al.,* 2014; *Bajamgnigni et al.,* 2017).

The direct evaluation method, called "ground truthing", is to compare various reanalysis and interpolated products against climate station observations for many points over a region of interest (*Wong et al.,* 2017; *Ebert et al.,* 2007). The indirect evaluation method uses different available climate forcing products separately to drive an H-LSM and then compares how well the various products simulate the streamflow (*Eum et al.,* 2014; *Bajamgnigni et al.,* 2017). The results of this indirect method, however, can be polluted by compensating effects, e.g., in parameterization. The indirect evaluation method is in general complementary to direct evaluation, because the streamflow at a gauge represents the integrated response of an upstream watershed, and its comparison against simulated streamflows describes the integrated quality of the climate forcing product used. The indirect evaluation approach is more practical if applied using default model parameterizations (without calibration), as then it will involve a reduced computational demand and also the results will not be complicated by the compensation effects of calibration.

The aim of this paper is to conduct a detailed analysis and evaluation of a physically-based H-LSM for a highly-managed, large-scale basin, using state-of-the-art calibration strategies and multiple data sources to enable quantification of modelling uncertainty. Such analysis is essential to comprehensively benchmark model performance, to examine water security vulnerabilities under future conditions, to serve as a test-bed (experimental basin) for improved model development, and to evaluate new datasets. Additionally, such analysis informs H-LSM applications for hydrologic operational forecasts and the management of large-scale basin water resources.

The specific objectives of this paper are as follows:

- To evaluate the quality of gridded precipitation datasets in terms of how well they reproduce observations of multiple streamflow gauges when used to drive an H-LSM.
- To improve the H-LSM parameterization using a state-of-the-art computationally-efficient calibration approach, and evaluate the effectiveness of parameter transferability through validation in time and space, using independent multiple streamflow gauges not used in calibration.
- To test the model performance using multiple sources of observational information on model storage and output fluxes, and to ensure that the optimal parameters obtained are as realistic as possible (giving the "right answers for the right reasons") without error compensation across multiple outputs.

The study area of this paper is the 406,000 km^2 Saskatchewan River Basin (SaskRB) located in western Canada (**Fig. 4.1**). The SaskRB is currently the focus of an extensive research initiative due to its vulnerability to potential climate change and extreme events such as droughts and floods (*Razavi et al.,* 2015; *Wheater and Gober,* 2015). The basin forms a complex system characterized by hydrologically distinct regions that include the Rocky Mountains, Boreal Forest, and the Prairies, all of which affect the regional and global hydroclimate in unique ways. The distinct hydrology-land surface processes are further complicated by extensive water management

(i.e., reservoir operations, diversions, and irrigation) and interjurisdictional water sharing policy (*Islam and Gan,* 2014; *Wheater and Gober,* 2015).

A limited number of hydrologic modelling studies have been conducted for different parts of the SaskRB. *Wen et al.* (2011) modelled agricultural drought using a long-term soil moisture over the Canadian Prairies using Variable Infiltration Capacity (VIC). VIC was calibrated using streamflow data and validated with observed soil moisture data. *Tanzeeba and Gan* (2011) used a Modified Interactions Soil-Biosphere-Atmosphere (MISBA) land surface scheme to study climate change impacts on some major subbasins of the SaskRB. MISBA was calibrated using naturalized streamflow at sub-basin outlets. Neither study accounted for reservoirs and irrigation, both of which significantly affect the simulation of flow and soil moisture. They also used only a limited number of streamflow stations for model calibration.

Recently, *Faramarzi et al.* (2016) applied the Soil and Water Assessment Tool (SWAT) model to assess model complexity and the effect of calibration strategy on freshwater estimation for Alberta watersheds, which includes the upstream subbasins of the SaskRB. The authors compared the basin-outlet calibration to multiple-station model calibration, while model complexities were changed by turning on and off some water management components. Their results showed that ignoring complex hydrological features resulted in inadequate model calibration and estimation of water resources. In their study, however, some of the cold region hydrological processes were represented in a simplified manner through conceptual models, such as the temperature-index approach for estimating snowmelt. Such simplistic process representation in cold regions may limit the credibility of the resulting watershed model, in particular under a changing climate. A more detailed energy-based approach is preferable to allow estimation of individual components of the energy balance to simulate the energy fluxes within the snowpack. An energy-based approach also allows direct estimation of evapotranspiration and sublimation and creates a stronger link to atmospheric modelling and remote sensing data (*Overgaard et al.*, 2006). The present study therefore examines the modelling of the SaskRB using an energy-based and

"process-based" H-LSM to calculate different hydrological components using fine-time resolution climate forcing data along with physically-based model parameterizations.

For the reliable modelling of the current and future hydrology of the SaskRB, comprehensive model development and testing are needed that consider the representation of water management and distinct hydrological processes occurring in the basin. For this purpose, a Canadian H-LSM, MESH (Modélisation Environmentale communautaire - Surface Hydrology) was used. MESH has been used extensively for Canadian watershed research studies (*Pietroniro et al.*, 2007; *Davison et al.*, 2006; *Mekonnen et al.*, 2014; *Haghnegahdar et al.*, 2014; *Yassin et al.*, 2017; *Mengistu and Spence,* 2016) and recent developments have included reservoir operations (*Yassin et al.,* 2019), irrigation, and flow diversion modules (the latter being incorporated for the first time in this study). This study is the first to report MESH model development for the entire SaskRB with representation of the aforementioned complexity and inclusion of detailed evaluation aimed to improve the understanding of the basin as a whole and create a test-bed for the simulation of alternative climate, land use and water management futures.

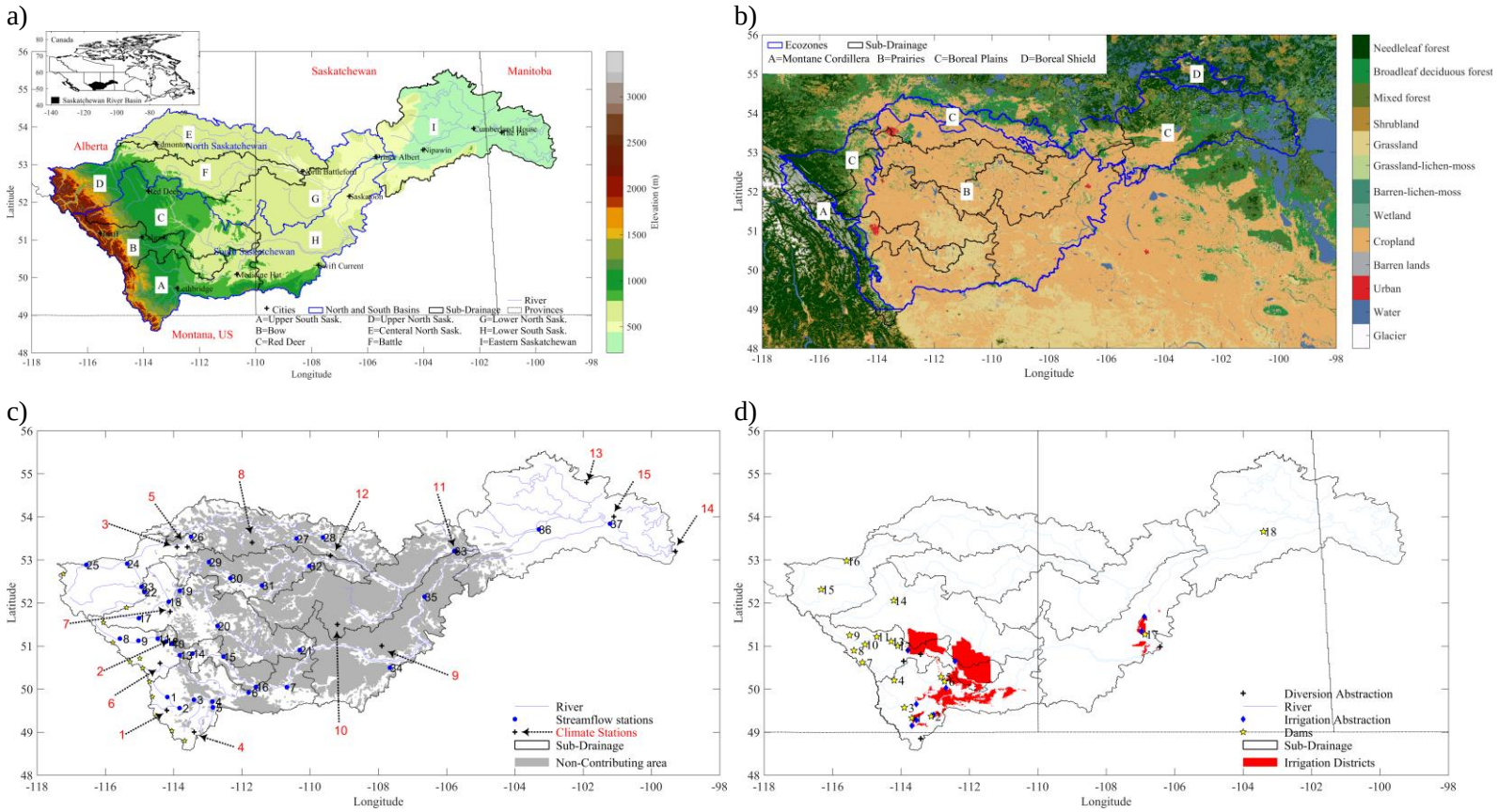


Figure 4. 1 Maps of study area, a) DEM, sub-drainages, cities, and river network, b) Land-cover map along with ecozones and sub-drainages of SaskRB, c) Streamflow stations, climate stations, and non-contributing map, d) Dams, irrigation districts, diversion, and irrigation abstraction points

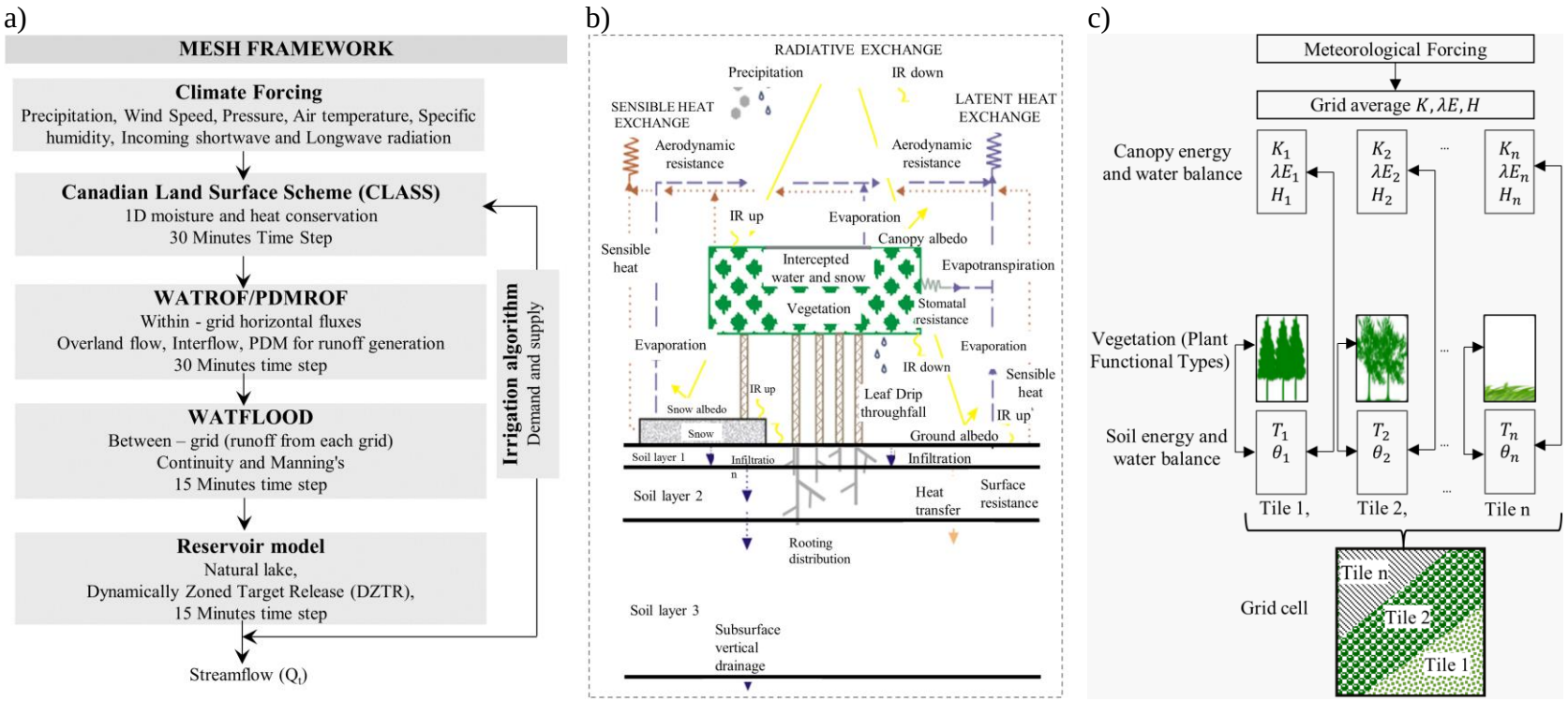


Figure 4. 2 a) MESH model schematic diagram and b) CLASS schematic diagram (*Verseghy* 2011), c) CLASS sub-grid structures for water and energy balance calculation (modified from (*Li and Arora*, 2012))

4.2 Study area

The Saskatchewan River Basin (SaskRB) (**Fig. 4.1a**) encompasses portions of the Canadian provinces of Alberta, Saskatchewan, and Manitoba, as well as a small portion of the US state of Montana. The SaskRB is situated in western Canada (98° - 118° W and 48° – 56° N), with a total drainage area of 406,000 km^2 and approximate maximum dimensions of 1300 km east-west and 700 km north-south. The source of the SaskRB originates from the eastern slopes of the Canadian Rockies in Alberta, which includes parts of the Columbia Icefield. The two main tributaries of the SaskRB are the South and North Saskatchewan Rivers, both of which flow east and northeast through Alberta and Saskatchewan Prairies, before merging to form the Saskatchewan River, flowing through the Saskatchewan Delta (North America's largest freshwater inland delta), and draining into Lake Winnipeg in Manitoba. The two tributary river systems are further subdivided into nine subbasins of the Bow, Red Deer, Battle, Upper North, Central North, Lower North, Upper South, Lower South, and Eastern Saskatchewan rivers. The Upper South basin can further be disaggregated into the Oldman basin and a small watershed draining from Montana, US.

The topographic elevation of the basin ranges between 218 and 3487 m above sea level (**Fig. 4.1a**). The physiographic characteristics extend from the rugged Canadian Rocky Mountains, foothills, and uplands on the far western side of the basin, to lowlands and plains in the remaining parts of the basin. The ecozones of the SaskRB (**Fig. 4.1b**) are classified into four ecozones; Montane Cordillera, Prairie, Boreal Plain, and Boreal Shield, covering 6%, 58%, 33% and 3% of the basin area, respectively. The Montane Cordillera Ecozone encompasses all the rugged mountains of the basin, the Boreal Plain has gently rolling to level topography, and the Boreal Shield contains hilly terrain with numerous ponds, wetlands, and lakes. The Prairie Ecozone covers post-glacial undulating plains to rolling plains and flat terrain with numerous depressional areas. The Prairies have several unique features: the pothole topography prevents some areas from draining to the major river system, the ecozone has internal drainage, and connections to the major river system are intermittent (*Pomeroy et al.*, 2010, *Shook et al.*, 2013). The parts of the ecozone not draining to the major river system are commonly called "non-contributing areas", defined as

the drainage areas not playing a part in runoff in a flood that has a two-year return period (*Godwin and Martin,* 1975). **Fig. 4.1c** shows the maximum possible ranges of the non-contributing areas in the SaskRB according to Prairie Farm Rehabilitation Administration (*PFRA, Hydrology Division,* 1983).

According to the Köppen climate classification, the SaskRB climate is classified as cold semi-arid and humid over the Prairies and cool and humid continental climate over the remaining parts, with a long, cold winter and a short, warm summer. The mean annual precipitation ranges from 601-800 mm in the Montane Cordillera to 201-600 mm in the Prairie and Boreal Ecozones. Most of the precipitation comes as rainfall between April and August, and snowfall occurs in winter between October and April. Much of the annual runoff is generated from snowmelt during early spring from the Prairies and during summer from the mountains.

Fig. 4.1b shows the land-cover of the SaskRB. The Montane Cordillera ecozone is covered by barren land, glacier, grassland, shrubland, as well as needleleaf, broadleaf and mixed forest. Cropland and grassland dominate the land cover of the Prairies followed by depressional wetlands and lakes. The Boreal ecozone around the upland and eastern part of the basin is covered by needleleaf and broadleaf forest, while the central flatland is largely covered by cropland with sparse broadleaf forest. The soil type in the Prairies is dominated by Chernozemic soils, clay-rich soils with a high water-holding capacity favorable for agriculture. The Boreal Plain soil types are Brunisol and Luvisol, productive for crop and tree growth. Both the eastern and a portion of the northwestern parts of the basin are underlain by mineral soils.

The SaskRB is regulated by many reservoirs and diversions for hydropower production, irrigation, and other water supply. Irrigation is the major consumptive demand in the basin. In the Alberta portion, there are 13 irrigation districts providing water to 1,412,836 acres of farmland. In the Saskatchewan portion, 11 districts (80,000 acres) receive water from Lake Diefenbaker. **Fig. 4.1d** shows the major reservoirs and irrigation districts included in this study. **Table 4.1** and **Table 4.2**, respectively, present brief summaries of the main reservoirs and irrigation districts.

Table 4. 1 Summary of major reservoirs in the Saskatchewan River Basin

No	Dam/Reservoir Name	Year	Main Purpose[1]	Long (°)	Lat (°)	Dam height (m)	Capacity (MCM[2])	C =Capacity/MAI[3]
1	St. Mary	1951	IR	-113.12	49.37	62	394.7	0.492
2	Waterton	1963	IR	-113.68	49.33	55	172.7	0.258
3	Oldman	1991	IR	-113.90	49.57	76	490.0	0.446
4	McGregor	1954	IR	-112.83	50.28	14	326.1	N/A
5	Travers	1954	IR	-112.72	50.18	41	317.0	N/A
6	Chain lakes	N/A	N/A	-114.162	50.267	N/A	17.3	N/A
7	Upper Kananaskis	1943	HP	-115.14	50.69	24	160.4	N/A
8	Spray Canyon	1951	HP	-115.37	50.89	60	421.9	N/A
9	Cascade	1942	HP	-115.50	51.25	35	387.3	N/A
10	Barrier lake	N/A	N/A	-115.060	51.004	N/A	24.0	N/A
11	Ghost	1929	HP	-114.71	51.22	42	132.0	0.048
12	Bearspaw	N/A	N/A	-114.30	51.14	N/A	17.0	N/A
13	Glenmore	1933	WS	-114.10	51.00	27	19.6	N/A
14	Dickson	1983	WS	-114.21	52.05	40	203.0	0.167
15	Big Horn	1972	HP	-116.33	52.31	150	1770.0	0.747
16	Brazeau	1962	HP	-115.59	52.97	89	490.0	N/A
17	Gardiner	1968	IR	-106.86	51.27	69	9870.0	1.460
18	E.B. Campbell	1962	HP	-103.40	53.66	34	2200.0	0.153

[1]Main purpose: WS-Water Supply, HP-Hydropower IR-Irrigation FC-Flood Control

[2] MCM – Million Cubic Meters

[3] MAI – Mean Annual Inflow

N/A Not Available

Table 4. 2 Summary of irrigation districts in Alberta (AB) and Saskatchewan (SK)

Name of Irrigation Districts	Province	Irrigation District Number	Year Establis-hed	Irrigated area (acres)	Source
Mountain View	AB	1	1931	1,052 (426)	Belly River
Leavitt	AB	1	1944	4,601(1,862)	Belly River
Aetna	AB	1	1959	1,929(781)	Belly River
United	AB	2	1923	17,277(6,992)	Belly and Waterton River
Magrath	AB	4	1900	11,188(4,528)	St. Mary, Waterton, Belly
Raymond	AB	4	1900	32,258(13,055)	St. Mary, Waterton, Belly
Lethbridge Northern	AB	3	1923	122,378(49,526)	Oldman River
Taber	AB	8	1917	76,872(31,110)	St. Mary, Waterton, Belly
St. Mary River	AB	4	1900	342,757(138,712)	St. Mary, Waterton, Belly
Ross Creek	AB	4	1954	1,055(427)	Ross Centre Creek
Bow River	AB	5	1920	198,196(80,209)	Bow River
Western	AB	7	1907	67,643(27,375)	Bow River
Eastern	AB	6	1914	274,940(111,267)	Bow River
Hillcrest	SK	9	1988	3,497	Lake Diefenbaker
South Saskatchewan River	SK	9	1966	38,349	Lake Diefenbaker
Macrorie	SK	9	1989	2,388	Lake Diefenbaker
Luck Lake	SK	8	1984	10,771	Lake Diefenbaker
Riverhurst	SK	8	1987	15,228	Lake Diefenbaker
Grainland	SK	8	1979	2,237	Lake Diefenbaker

4.3 Data

4.3.1 Model setup data

Topographic data were acquired from Canadian Digital Elevation (CDED, 2013) https://open.canada.ca/data/ at a scale of 1:250,000. The land-cover map was extracted from the Canada Centre for Remote Sensing (CCRS) (North American Land Cover, 2005), which resulted in 14 land-cover types for the SaskRB (**Fig. 4.1b**). The gridded soil texture data on percentage sand, clay, and organic matter from the Regridded Harmonized World Soil Database v. 1.2 (*Wieder et al.,* 2014), were re-gridded and upscaled from the Harmonized World Soil Database (HWSD) v. 1.2 (*Nachtergaele et al.,* 2009). Historical hydrometric data for model evaluation and calibration stations shown in **Fig. 4.1c** were obtained from the National Water Data Archive of the Water Survey of Canada. The list and brief details of the hydrometric stations are presented in **Table 4.3**. The irrigated cropland fraction was estimated using the Global Map of Irrigation Areas (GMIA) (*Siebert et al.,* 2013). The data required for reservoir operation schemes such as maximum capacity, reservoir area-level-volume, time series of outflow-storage-inflow and environmental release were obtained from the Water Survey of Canada, Saskatchewan Water Security Agency, and Alberta Environment and Parks.

Table 4. 3 Streamflow stations for calibration and validation model (Figure 1c)

	Station	Station Name	Province[1]	Latitude (°)	Longitude (°)	Drainage Area (km^2)	Regulated	Subbasin	Operation Schedule
1	05AA023	Oldman River near Waldrons Corner	AB	49.81	-114.18	1446	False	A	Continuous
2	05AA024	Oldman River near Brocket	AB	49.55	-113.82	4400	True	A	Continuous
3	05AB046	Willow Creek at Highway NO. 811	AB	49.75	-113.40	2510	False	A	Seasonal
4	05AD007	Oldman River near Lethbridge	AB	49.71	-112.86	17000	True	A	Continuous
5	05AE006	St. Mary River near Lethbridge	AB	49.57	-112.84	3530	True	A	Continuous
6	05AG006	Oldman River near the Mouth	AB	49.91	-111.80	27500	True	A	Continuous
7	05AJ001	South Saskatchewan River at Medicine Hat	AB	50.04	-110.67	56400	True	A	Continuous
8	05BB001	Bow River at Banff	AB	51.17	-115.57	2210	False	B	Continuous
9	05BE004	Bow River near Seebe	AB	51.12	-115.00	5170	True	B	Continuous
10	05BH004	Bow River at Calgary	AB	51.05	-114.05	7870	True	B	Continuous
11	05BH005	Bow River near Cochrane	AB	51.17	-114.46	7410	True	B	Seasonal
12	05BH008	Bow River Blow Bearspaw Dam	AB	51.09	-114.22	7770	True	B	Continuous
13	05BL024	Highwood River near the Mouth	AB	50.78	-113.82	3950	True	B	Continuous
14	05BM002	Bow River Below Carseland Dam	AB	50.82	-113.44	15700	True	B	Seasonal
15	05BM004	Bow River Below Bassano Dam	AB	50.75	-112.54	20300	True	B	Seasonal
16	05BN012	Bow River near the Mouth	AB	50.04	-111.59	25300	True	B	Continuous
17	05CA009	Red Deer River below Burnt Timber Creek	AB	51.64	-115.01	2250	False	C	Continuous
18	05CB001	Little Red Deer River near the Mouth	AB	52.02	-114.14	2580	False	C	Continuous
19	05CC002	Red Deer River at Red Deer	AB	52.27	-113.81	11600	True	C	Continuous
20	05CE001	Red Deer River at Drumheller	AB	51.46	-112.71	24900	True	C	Continuous
21	05CK004	Red Deer River near Bindloss	AB	50.90	-110.29	47800	True	C	Continuous
22	05DB006	Clearwater River near Dovercourt	AB	52.25	-114.85	2250	False	D	Continuous
23	05DC001	North Saskatchewan River near Rocky Mountain House	AB	52.37	-114.94	11000	True	D	Continuous
24	05DD005	Brazeau River Below Brazeau Plant	AB	52.91	-115.36	5660	True	D	Continuous
25	05DD007	Brazeau River Below Cardinal River	AB	52.88	-116.55	2600	False	D	Seasonal
26	05DF001	North Saskatchewan River at Edmonton	AB	53.53	-113.48	28100	True	D	Continuous
27	05EE007	Vermilion River near Marwayne	AB	53.49	-110.39	7260	True	E	Seasonal
28	05EF001	North Saskatchewan River near Deer Creek	SK	53.52	-109.61	57200	True	E	Continuous
29	05FA011	Battle River at Duhamel	AB	52.94	-112.96	5010	False	F	Continuous
30	05FC001	Battle River Near Forestburg	AB	52.57	-112.34	7680	True	F	Seasonal
31	05FC008	Battle River At Highway No. 872	AB	52.40	-111.41	11700	True	F	Seasonal
32	05FE004	Battle River near the Saskatchewan Boundary	AB	52.85	-110.01	25100	True	F	Continuous
33	05GG001	North Saskatchewan River at Prince Albert	SK	53.20	-105.77	131000	True	G	Continuous
34	05HD039	Swift Current Creek near Leinan	SK	50.49	-107.65	3730	True	H	Continuous
35	05HG001	South Saskatchewan River at Saskatoon	SK	52.14	-106.64	141000	True	H	Continuous
36	05KD003	Saskatchewan River below Tobin Lake	SK	53.70	-103.29	289000	True	K	Continuous
37	05KJ001	Saskatchewan River at the Pas	MB	53.83	-101.20	389000	True	K	Continuous

1 Alberta (AB), Saskatchewan (SK), and Manitoba (MB)

4.3.2 Climate datasets

We chose five gridded precipitation products for evaluation of precipitation uncertainty. These datasets were selected to meet four main criteria. The first criterion is the sub-hourly temporal resolution that is needed to run the land surface model. The second criterion is the spatial resolution needed to cover the study area with grid resolution $\leq 0.5°$. The third criterion is the dataset must provide complete data for the simulation period intended. The last criterion is that a dataset must be a blended product from multiple sources to address limitations inherent in single-source precipitation products (e.g. station-based, satellite-derived) (*Xie and Arkin,* 1996; *Maggioni et al.,* 2014; *Shen et al.,* 2010). Furthermore, products that are based on similar integration methods are not considered. As a result, other commonly and widely used datasets are excluded. The examples of these include the monthly Canadian Gridded temperature and precipitation (CANGRD) dataset (*Zhang et al.,* 2000), the daily station-interpolated ANUSPLIN, the coarser resolution Japan Meteorological Agency 55-year Reanalysis (JRA-55) (*Onogi et al.,* 2007; *Kobayashi et al.,* 2015) and the Modern-Era Retrospective Analysis for Research and Applications (MERRA) (*Rienecker et al.,* 2011), numerous satellite-derived products such as TRMM (*Huffman et al.,* 2002), TMPA (*Huffman et al.,* 2007), PERSIANN-CCS (*Hsu et al.,* 2010), and CMOPRH (*Joyce et al.,* 2004), and ERA-40 (*Uppala et al.,* 2005) as an older version of ERA-Interim. In addition, some products (e.g., the Multi-Source Weighted-Ensemble Precipitation (MSWEP) (*Beck et al.,* 2017) and ERA-5) that fulfill the above criteria have been left out because they became available after the accomplishment of this study. A brief description and summaries of these five gridded precipitation products is provided in the following and **Table 4.4.**

The Canadian Precipitation Analysis (CaPA) was created to provide a dataset of 6-hourly precipitation accumulations over North America from 2002 onwards at a spatial resolution of 15 km (*Mahfouf et al.,* 2007). For Canada, the regional Global Environmental Multiscale (GEM) model was used to generate data by an optimum interpolation technique, in which the initial guess from the model was updated by the rain-gauge measurements. CaPA has been continuously

improving, assimilating data from the Canadian weather radar network and US radars near the border, and recently, refining its spatial resolution to 10 km (*Fortin et al.,* 2015).

The Terrestrial Hydrology Research Group at the Princeton University had produced a global dataset of 3-hourly meteorological data at a spatial resolution of 1.0° spatial resolution (~120 km) from 1948 to 2000 (*Sheffield et al.,* 2006). The components in generating this dataset (called hereafter "Princeton") included the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis and a set of global observation-based data. Princeton has been updated and the 1901-2012 version at 0.5° and 3-hourly time steps was used in this study.

The European Union Water and Global Change (WATCH) Forcing Data methodology applied to the ERA-Interim (WFDEI) was developed to provide global datasets of sub-daily (3-hourly) and daily meteorological data at a spatial resolution of 0.5° (~50 km) covering the period of 1979 to 2012 (*Weedon et al.,* 2014). Similar to Princeton, WFDEI was constructed based on the European Centre for Medium-Range Weather Forecasts Re-Analysis Interim product, combined with the Global Precipitation Climatology Centre (GPCC) monthly variables and the climatic Research Unit (CRU) monthly data. Therefore, WFDEI has used either GPCC or CRU precipitation totals to produce two sets of rainfall and snowfall data. In this study, both sets of data were used, for brevity, called hereafter GPCC and CRU.

The North American Regional Reanalysis product (NARR) was created to provide 3-hourly meteorological data for the North America domain at a spatial resolution of 32 km (~0.3°) from 1979 to 2015 (*Mesinger et al.,* 2006). NARR was derived from the NCEP-Department of Energy (NCEP-DOE) reanalysis and was combined with the NCEP regional Eta model, the Noah land-surface model, and numerous additional data sources. The assimilation of station observations over Canada has been discontinued since 2004 onwards (*Mesinger et al.,* 2006).

To drive the MESH model and to evaluate precipitation uncertainty, the above precipitation datasets were individually combined with the other six required forcing variables (wind speed, air

temperature, incoming shortwave radiation, specific humidity, incoming longwave radiation, and barometric pressure) gathered from the Global Environmental Multiscale (GEM) numerical weather prediction model (*Côté et al.*, 1998). It is recognized that there will be some inconsistencies with the precipitation fields from the other variables.

Table 4. 4 Precipitation products used for comparison

Dataset	**Full name**	**Type**	**Spatial resolution**	**Temopral resolution**	**Duration**	**Coverage**	**Reference**
CaPA	Canadian Precipitation Analysis	Station-based multiple source	300 arcsec (~0.0833°) /~10 km)	6h	2002-2014	North America	Mahfouf *et al.* (2007)
Princeton	Global dataset at the Princeton University	Reanalysis-based multiple-source	0.5°(~50km)	3h	1901-2012	Global	Sheffield *et al.* (2006)
WFDEI[CRU]	Water and Global change Forcing Data methodology applied to ERA-Interim [Climate Research Unit]	Reanalysis-based multiple-source	0.5°(~50km)	3h	1979-2012	Global	Weedon *et al.* (2014)
WFDEI[GPCC]	Water and Global Change Forcing Data mehodlolgy applied to ERA-Interim [Global precipitation Climatology Center]	Reanalysis-based multiple-source	0.5°(~50km)	3h	1979-2012	Global	Weedon *et al.* (2014)
NARR	North American Regional Reanalysis	Reanalysis-based multiple-source	32 km (0.3°)	3h	1979-2015	North America	Mesinger *et al.* (2006)

4.4 Description of MESH modelling system

4.4.1 MESH core components

MESH is Environment and Climate Change Canada's (ECCC) H-LSM framework (*Pietroniro et al.,* 2007) encompassing several types of modelling structure (**Fig. 4.2a**). MESH uses an evenly gridded spatial organization approach to configure the landscape. Sub-grid heterogeneity is represented by dividing each grid based on tiles defined by land-cover classes (in this study each type represents one tile) or based on other user-specified mosaic options. Each land-cover class has similar hydrological responses. MESH runs in an off-line mode at a half-hourly time step using seven meteorological forcing data variables at 40 m height. These are precipitation (mm s^{-1}), air temperature (K), wind speed (m s^{-1}), incoming shortwave radiation (W m^{-2}), incoming longwave radiation (W m^{-2}), specific humidity (kg kg^{-1}) and barometric pressure (Pa).

MESH contains vertical, lateral (within grid) and grid-to-grid routing components. In the vertical, water and energy balances are calculated at tile scale with the Canadian Land Surface Scheme (CLASS) (**Fig. 4.2b**) (*Verseghy,* 1991, 2000; *Verseghy et al.,* 1993). The CLASS basic prognostic variables of CLASS include the soil layers temperatures, the soil layers liquid and frozen moisture contents; the mass, temperature, density, albedo and liquid water content of the snow pack; the temperature of the vegetation canopy and the mass of intercepted rain and snow present on it; the temperature and depth of ponded water on the soil surface; and an empirical vegetation growth index (*Verseghy* 2011). CLASS preserves the prognostic variables of each of the tiles between time steps, while the surface fluxes are averaged using the tiles' fractional weight in each grid cell (**Fig. 4.2c**). CLASS contains four plant functional types, including needleleaf forest, broadleaf forest, crop, and grass; other vegetation units are approximated to one of the plant functional types by adjusting the parameter values. Glaciers are represented as a one-dimensional ice column. Water and energy balances are computed on the ice sheet using ice volumetric heat capacity and thermal conductivity. The default configuration of the CLASS soil layer contains three layers with a thickness of 0.10, 0.25, and 3.75 m, respectively. The Green-Ampt method is

used to estimate infiltration rate through the soil profile, and soil water storage and transmission to gravitational and soil moisture suction forces is calculated using a 1-D Richards' equation. The soil hydraulic properties are estimated using gridded soil texture data. The snowmelt process is governed by an energy budget approach. Evapotranspiration is estimated using a bulk mass transfer equation dependent on humidity (vapor pressure gradient). Interception of water and snow by vegetation is calculated as a function of plant leaf area index.

The lateral processes in MESH include: 1) blowing snow across tiles within a grid square, 2) lateral water movement in the soil, and 3) excess surface water flow from the tiles to the drainage system. Blowing snow transport and sublimation quantities are calculated within the grid square across tiles using the Prairie Blowing Snow Model (PBSM) (*MacDonald et al.,* 2009). Wind-eroded snow from a tile either sublimates or transports and deposits into downwind tiles (according to aerodynamic roughness or drifting in a descending order) in the same grid square (*MacDonald et al.,* 2009). Lateral movement of water in the soil and water on the surface is computed with either of the algorithms PDMROF (*Mekonnen et al.,* 2014) or WATROF (*Soulis, et al.,* 2000). WATROF has been introduced to enhance the hillslope hydrology representation in MESH (*Soulis, et al.,* 2000). Lateral flow in the soil is simulated as a function of lateral flow for unsaturated and saturated condition via the bulk value of the soil layer moisture by means of an approximated Richards' equation (*Soulis et al.,* 2000, 2011). Surface overland flow is routed from the tile surface to drainage network within a model grid cell using Manning's approximation of the kinematic wave velocity equation. PDMROF has been introduced to calculate the variable contributing nature of Prairie regions. PDMROF is designed based on the probability density model (PDM) model of *Moore* (2007) to parsimoniously characterize the Prairies' dynamic contributing area behaviour (*Mekonnen et al.,* 2014).

The third lateral process, the routing component, includes grid-to-grid river flow routing and reservoir operation. It uses the WATFLOOD routing algorithm (*Kouwen et al.,* 1993) to route the flows collected from overland flow, interflow and drainage through a watershed using a storage routing technique in which the outflow discharge is estimated as a function of water storage in the

channel, computed using the continuity and Manning's equations. MESH reservoir operation along with irrigation and diversion is discussed in the next section.

4.4.2 MESH new features

To enable MESH to model complex and highly managed basins (e.g. SaskRB), new water management features (irrigation, reservoir operation, and diversion) have been integrated recently into the MESH framework.

The irrigation algorithm is based on the soil moisture deficit approach, similar to that of *Pokhrel et al.* (2016). The net irrigation water demand is estimated as the difference between target soil moisture content (θ_T) and the simulated soil moisture (θ_k).

$$IR = \frac{\rho_w}{\Delta t}\sum_{k=1}^{n}\{\max[(\theta_T - \theta_k), 0] \times D_k\} \quad (1)$$

where IR [kg m^{-2} s^{-1}] is the net irrigation demand, ρ_w[kg m^{-3}] is the density of water; Δt is model time step; θ_T is given as $\alpha \times \theta_{FC}$; θ_{FC} and θ_k [m^3 m^{-3}] are the field capacity and simulated actual volumetric soil moisture content, respectively; α [-] is the parameter that defines the upper soil moisture limit that has been used varyingly from 0.5 to 1; and D_k [m] is the thickness of κ^{th} soil layer, *n* represents the number of soil layers considered in the calculation. To represent irrigation effects, the standard CLASS three soil layer configuration has been changed to four soil layers so that the bottom of the third soil layer is set to around 1 m. The thickness of each soil layer is 0.1, 0.25, 0.7, and 3.05 m. The top three layers are considered for irrigation with a crop root depth of 1.0 m. The estimated irrigation demand is applied to the soil as rain between 0600 and 1000 local time each day in a similar approach as *Ozdogan et al.,* (2010) and *Pokhrel et al.,* (2016). The excess irrigation water (return flow) is assumed to join the nearest river system in the form of interflow and bottom-layer soil drainage.

Reservoir regulation is represented by the Dynamically Zoned Target Release (DZTR) scheme that uses a parametric piecewise-linear function to approximate actual reservoir release rules (*Yassin et al.,* 2019a). The DZTR scheme divides reservoir storage into five zones, dead storage (Zone 0), critical storage (Zone 1), normal storage (Zone 2), flood storage (Zone 3) and

emergency storage (Zone 4). Whenever storage is below full supply storage zone, the release only occurs at the bottom outlet, but when storage is within flood storage, the release happens from both outflow outlet and spillway. The dead storage (Zone 0) amount is assumed 10% of the maximum storage or a dead storage value from the reservoir characteristics data. In general, where no operational information is available, the other storage zones are estimated from historical time series of storage by defining some non-exceedance probability value for each zone or by optimizing these zones to reproduce the observed storage and release time series. Target releases for each zone are obtained in a similar fashion. These target storages and releases are allowed to vary each month (or on any other arbitrarily selected time step) to allow a better representation of the seasonality of reservoir operation.

Zone 0 $Q_t = 0$ $[S_t < 0.1S_{max}]$ (2)

Zone 1 $Q_t = \min\left(Q_{ci}, \frac{S_t - 0.1S_{max}}{\Delta t}\right)$ $[0.1S_{max} < S_t \leq S_{ci}]$ (3)

Zone 2 $Q_t = Q_{ci} + (Q_{ni} - Q_{ci})\frac{(S_t - S_{ci})}{(S_{ni} - S_{ci})}$ $[S_{ci} < S_t \leq S_{ni}]$ (4)

Zone 3A $Q_t = Q_{ni} + (Q_{mi} - Q_{ni})\frac{(S_t - S_{ni})}{(S_{mi} - S_{ni})}$ $[S_{ni} < S_t \leq S_{mi}]$ (5A)

Zone 3B $Q_t = Q_{ni} + \max\{(I_t - Q_{ni}), (Q_{mi} - Q_{ni})\}\frac{(S_t - S_{ni})}{(S_{mi} - S_{ni})}$ $[S_{ni} < S_t \leq S_{mi}]$ (5B)

Zone 4 $Q_t = \min([\max\left(\frac{(S_t - S_{mi})}{\Delta t}, Q_{mi}\right)], Q_{mc})$ $[S_{mi} < S_t]$ (6)

where I_t, Q_t and S_t are inflow, release and storage at time step t. S_{ci} , S_{ni} and S_{mi} are critical, normal and maximum storage targets for month i. Q_{ci} , Q_{ni} and Q_{mi} are critical, normal and maximum release targets for month i. Q_{mc} is maximum channel capacity parameter.

We also developed a flow diversion process within MESH to represent water transfer across the basin via engineered works, for various purposes including irrigation. Flow diversion is the water transfers within-basin from one river node to another, water transfers from outside a basin to a within-basin river node, and water withdrawals from river node to irrigated areas. The

flow diversion implementation in MESH is divided into two types depending on the location of source and sink of diverted water: type 1 has either a water source or sink located outside the whole basin; type 2 has both sources and sinks located within the basin. To divert water from one point to another, the locations of sources and sinks and the amount to be diverted at each time step are provided as input to the model. However, in the case of flow diversion for irrigation, the amount of water for diversion is estimated using the irrigation demand algorithm discussed above.

4.5 Modelling methodology

4.5.1 MESH model configuration

A grid resolution of 0.125° was used to configure MESH model which resulted in 3,667 grid cells for the SaskRB. Gridded watershed characteristics are derived based on the topographic data defined above (**Fig. 4.1a**). The flow direction and watershed delineation was analysed at a terrain data native resolution of 1:50,000, and then the GreenKenue tool (*Canadian Hydraulics Centre*, 2010) was used to upscale to 0.125° modelling grids. The HWSD silt, sand, and clay percentage has a spatial resolution of 0.05°. The percentage of soil texture and land cover were regridded to 0.125° from their native resolution using area weights. Gridded irrigated cropland fraction was estimated using GMIA irrigation fraction and CCRS cropland fraction. The CCRS cropland was separated into irrigated and non-irrigated cropland. The irrigation district boundaries (**Fig. 4.1d**) and model setup grids have been used as an intersecting layer while extracting the irrigated fraction GMIA. The data and parametrization of the DZTR reservoir operation scheme were obtained from *Yassin et al.* (2019), using storage, inflow and release daily time series for each reservoir. The non-contributing areas of the SaskRB (**Fig. 4.1c**) delineated by the Prairie Farm Rehabilitation Administration (*PFRA, Hydrology Division,* 1983) were used to configure grids as contributing and non-contributing. PDMROF (*Mekonnen et al.,* 2014) and WATROF (*Solis et al.,* 2000) were used to drive the soil and surface water lateral movement for regions identified as non-contributing areas and contributing areas, respectively.

4.5.2 Evaluation of precipitation data uncertainty and model performance

The precipitation and streamflow performance was evaluated using nine performance metrics that measure the goodness of fit on precipitation, different components of the hydrograph, evapotranspiration (ET), and total water storage anomaly (TWS). The equations for the nine metrics are given in **Table 4.5**, where (1) PBIAS(FL) is the percentage of flow volume bias between simulated and observed flows, (2) NSE(FL) Nash Sutcliffe Efficiency on the streamflow, (3) NSE(logFL) is Nash Sutcliffe Efficiency on the logarithm of streamflow to put emphasis on fitting low flows, (4 and 5) r(TWS) and r(ET) are the correlation coefficients for TWS anomaly and for evapotranspiration (ET), measuring the agreement between observed and simulated TWS and ET, respectively. The value of NSE (FL) and NSE (logFL) ranges between $-\infty$ and 1.0 with an optimal value of one. The value of PBIAS (FL) varies between -100 % to $+\infty$ with an optimal value of 0 %. A negative PBIAS (FL) indicate volume underestimation, and positive values indicate volume overestimation.

Table 4. 5 Precipitation and streamflow performance metrics

Streamflow performance metrics			
Performance Measure	**Symbol**	**Equation**	
Percentage of Bias	PBIAS(FL)	$\frac{\sum(Q_{obs} - Q_{sim})}{\sum Q_{obs}} * 100$	(1)
Nash Sutcliffe Efficiency	NSE(FL)	$1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2}$	(2)
Nash Sutcliffe Efficiency on log flow	NSE(logFL)	$1 - \frac{\sum\left(ln(Q_{obs}) - ln(Q_{sim})\right)^2}{\sum\left(ln(Q_{obs}) - ln(\overline{Q_{obs}})\right)^2}$	(3)
Correlation coefficients on TWS	r(TWS)	$\frac{\sum_i^N (TWS_{sim} - \overline{TWS_{sim}})\,(TWS_{obs} - \overline{TWS_{obs}})}{\sqrt{\sum_i^N (TWS_{sim} - \overline{TWS_{sim}})^2}\sqrt{\sum_i^N (TWS_{obs} - \overline{TWS_{obs}})^2}}$	(4)
Correlation coefficients on ET	r(ET)	$\frac{\sum_i^N (ET_{sim} - \overline{ET_{sim}})\,(ET_{obs} - \overline{ET_{obs}})}{\sqrt{\sum_i^N (ET_{sim} - \overline{ET_{sim}})^2}\sqrt{\sum_i^N (ET_{obs} - \overline{ET_{obs}})^2}}$	(5)
Q_{obs}: Observed flow, Q_{sim}: Simulated flow, $\overline{Q_{obs}}$: Mean of observed flows			

4.5.3 Model-criterion optimization

Model calibration was conducted using an aggregated multi-objective optimization approach. The aggregation uses equal weights to combine three objective functions defined on streamflow PBIAS (FL), NSE (logFL) and NSE (FL), as shown in Eq. 7. To attain a spatially-consistent model performance, each objective function is evaluated on multiple calibration streamflow stations individually, and then averaged to provide a basin-wide performance for use in calibration; this approach is commonly referred to as the "global calibration" approach. Global calibration is regarded as a reasonable trade-off between local performance and regional consistency of parametric information (*Haghnegahdar et al.*, 2014; *Ricard et al.*, 2013).

$$\underset{x\in\theta}{\text{minimize}}\, F(x) = \min_{x\in\theta}\left\{\left[\frac{\sum_i^m \text{abs(PBIAS(FL)}(x)_i)}{m}\right] + \left[\frac{\sum_i^m \text{-1}\times\text{NSE(logFL)}(x)_i}{m}\right] + \left[\frac{\sum_i^m \text{-1}\times\text{NSE(FL)}(x)_i}{m}\right]\right\} \qquad 7$$

θ: Parameter space, x: model parameters (decision variables), m: total number of streamflow stations averaged

The number of parameters that require calibration in MESH increases with the number of land-cover classes, complicating the parameter identification and equifinality challenges. Calibrating a large number of soil, vegetation and snow parameters only based on streamflows can lead to significant errors in other model state variables and other flux outputs (*Yassin et al.*, 2017). Improved calibration can be achieved by using multiobjective-multivariate calibration with a large number of model evaluations. Running MESH for the SaskRB for seven years in series requires ~10 hours, and to complete 5,000 model evaluations on a parallel computing system with 10 nodes with 16 cores per node needs around two weeks of computational time. Such a high computational demand creates difficulty to conduct multivariate calibration that requires more than 20,000 model evaluations.

Alternatively, this study calibrated a sub-set of parameters chosen using the previous sensitivity analysis results on MESH (*Yassin et al.*, 2017; *Haghnegahdar et al.*, 2014; 2017) (**Table 4.6**). The selected sub-set includes parameters of river routing, some of the CLASS vegetation parameters, and the conceptual parameters of PDMROF and WATROF. These parameters were shown to control most of the variability in the model performance (*Yassin et al.*, 2017; *Haghnegahdar et al.*, 2017) and are shortlisted here as essential parameters for calibration, while other 'less influential' parameters were set at their default values. This reduction in the calibration problem helped us reduce the associated computational burden, while obtaining consistent and well-performing values for the parameters that dominantly control the model behavior. The results from a model run using default parameters (a priori) are used to benchmark model performance improvement through calibration. The parameter values and bounds (**Table 4.6**) in both configurations were specified based on suggested values from the CLASS technical manual (*Verseghy* 2011), previous studies and other H-LSMs suggested values.

The calibration of the selected model parameters was conducted using the Dynamically Dimensioned Search (DDS) method (*Tolson and Shoemaker* 2007). The DDS algorithm has been demonstrated to have advantages over other optimization approaches when the model is computationally intensive, because it requires a relatively low number of model evaluations

(<10,000) to achieve a good global solution (*Tolson and Shoemaker* 2007). This is a very important factor and hence the algorithm is well suited for the MESH model, which is computationally intensive; applying an ideal number of model evaluations for convergence of the optimal solution or Pareto front for MESH would be extremely time consuming particularly when the model is run on a large watershed (e.g. the SaskRB) with a large number of model parameters (decision variables).

Table 4. 6 Parameters and their corresponding ranges for calibration of the MESH model

Parameter	Description	Range
PDMROF parameters		
CMAX	Maximum storage parameter [m]	(0.01, 5)^{C,IC,G}
B	Shape factor parameter []	(0.01, 10) ^{C,IC,G}
WATROF parameters		
MANN	Manning's roughness coefficient 'n'	(0.05, 0.16) $^{NF,BF,MF, SL}$ (0.05, 0.16) ^{G,GL,C,IC,B,BL}
KSAT	Saturated surface soil conductivity (m s^{-1})	(0.00001, 0.10) $^{NF,BF,MF, SL}$ (0.00001, 0.10) ^{G,GL,C,IC,B,BL}
River routing and baseflow parameters		
R2N	Channel Manning's rougness (N=9)	(0.03, 0.16)
R1N	Overbank Manning's rougness (N=9)	(0.03, 0.16)
LZF	Constant on lower zone function (N=9)	(0.000001, 0.0001)
PWR	Exponent on the lower zone storage (N=9)	(1.00, 3.00)
CLASS parameters		
LAMAX	Annual maximum leaf area index	(3.00,10.00) ^{BF} (3.00,5.00) $^{CC, IC}$ (3.00,8.00)^{SL} (3.00,5.00)^{GG,GL} (0.50,3.00) ^{NF}
LNZO	Natural logarithm of the roughness length	(0.00,1.10) ^{BF} (-2.53,-1.05) $^{CC, IC}$ (0.00,1.10)^{SL} (-3.91,-2.53)^{GG,GL} (0.00,0.69) ^{NF} (-4.60,-3.90)^{BB,BL}
ALVC	Average visible albedo when fully leafed	(0.02,0.10) ^{BF} (0.02,0.10) $^{CC, IC}$ (0.02,0.10)^{SL} (0.02,0.10)^{GG,GL} (0.02,0.10) ^{NF} (0.02,0.10)^{BB,BL}
ALIC	Average near-infrared albedo when fully leafed	(0.20,0.40) ^{BF} (0.20,0.40) $^{CC, IC}$ (0.20,0.40)^{SL} (0.20,0.40)^{GG,GL} (0.20,0.40) ^{NF} (0.20,0.40)^{BB,BL}
SDEP	Soil permeable depth (m)	(0.7, 4.1) $^{NF,BF,MF, SL,GL,BB,BL}$
Ranges for different land-cover types: NF= Needleleaf Forest, BF=Broadleaf Forest, MF= Mixed Forest, SL=Shrubland, G=Grassland, GL=Grassland lichen moss, B=Barrenland, BL=Barren lichen moss, C=Cropland, IC, Irrigated Cropland, N number of classification over the basin		

4.5.4 Model calibration and verification configuration

Model calibration and validation were carried out using data from a 10-year period. An initialization test was conducted to determine the required warm-up period, and the results showed (not reported here) that a one year warm-up brings the prognostic variables close to stability. The first year (2002) was used for model warm-up, the next six years (2003-2008) for model calibration and the last three years (2009-2011) for model validation. Thirty-seven streamflow stations available through the Water Survey of Canada for SaskRB have been used for calibration and validation purposes (**Fig. 4.1d** and **Table 4.3**). These stations were screened based on two criteria, (1) having drainage areas > 1500 km^2, and (2) continuous streamflow data record lengths at least for the model calibration period (2003-2008). Twenty-two out of thirty-seven stations (60%) have been used for model temporal calibration and validation, while the remaining 15 (40%) stations were used for spatial validation (independent stations for validation). Drainage areas of the stations used for calibration ranged from approximately 1500 to 289,000 km^2, and of those used for spatial validation ranged from 2580 to 389,000 km^2.

As an additional way of model validation, the simulated total water storage anomaly (TWS) and evapotranspiration (ET) were evaluated against corresponding available observations. The modelled TWS were compared against TWS anomaly observation by GRACE satellite as validation. For this study, the GRACE TWS anomaly data were obtained from Natural Resource Canada in which the GRACE TWS data is estimated by means of a two-step filtering approach (*Huang et al.,* 2012; *Lambert et al.,* 2013) at 1°×1° resolution for the period of 2003-2011. The TWS anomaly of the GRACE is relative to the mean of January to December 2009.

Furthermore, we evaluated the simulated evapotranspiration using two observed datasets: (1) global NDVI-based gridded monthly evapotranspiration data (*Zhang et al.,* 2010) in which NDVI is used with surface resistance to estimate transpiration and soil evaporation using a modified Penman-Monteith method and open water evaporation using the Priestley-Taylor approach. These data are available from 1983 and more details about the estimation algorithm are provided in *Zhang et al.* (2010), (2) ET estimated based on latent heat flux observations for two

flux towers, namely, “Old Jack Pine (OJP)” and “Old Black Spruce (OBS)” located in the Boreal plain of the SaskRB near the basin outlet. The flux tower observations are available for the period of 1999-2009.

4.6 Results and discussion

The results are presented here in four sections. In section 4.6.1, the performance of each precipitation data on seasonal streamflow simulation are discussed. Section 4.6.2 presents the model calibration and validation results. Lastly, section 4.6.3 display the performance of calibrated model on other flux and store outputs as additional validation.

4.6.1 Precipitation data intercomparison based on streamflow simulations

The streamflow performance metrics for different precipitation data were compared by keeping model configuration and parameterizations the same; the only change factor was the precipitation data. The comparison based on streamflow represents the integrated response of an upstream watershed to describe the integrated effect of precipitation data quality. The streamflow simulation comparison was evaluated without calibration using default (a priori) parameter values for two main reasons: 1) To minimize the effect of mixing with and compensating for other errors arising from the process representation, model structural uncertainties, and parameter uncertainties; and 2) to evaluate the performance of different precipitation datasets on streamflow simulations before calibrating the model with all precipitation products, which is very computational demanding. Evaluating the datasets first enabled us to select the best performing precipitation data for subsequent calibration.

Regarding the predictive power of the model (**Fig. 4.3a**), in the Upper North Saskatchewan River basin, the performance of the headwater stream gauges (station 1, 2, and 5) were the lowest (0.04, 0.19, and 0.18) across the basin. The model performance gradually improved when going downstream with NSE values increased from 0.39 (station 3) to 0.45 (station 6). In particular, GPCC and CaPA performed better than the other precipitation datasets in spring and summer. The model performed similarly in the Bow River basin where the NSE (FL) values of the headwater

streamflow gauges (stations 9 to 12) were below 0.4, and those of the downstream gauges (stations 13 to 16) were above 0.6. The poor performance in the Lower South Saskatchewan River basin (NSE values all below zero in all seasons except station 35 in autumn) was also probably due to the direct impact of the high magnitude of error in the precipitation products (*Wong et al.,* 2017). For the other subbasins in the SaskRB, GPCC and CaPA generated the best streamflow performance in all seasons except winter. CaPA resulted in the best model performance, with the grand NSE (FL) of 0.35. However, Princeton had the overall lowest NSE (FL) at most stations and seasons, with the grand NSE (FL) (-0.81). The precipitation comparison based on NSE (logFL) showed a similar pattern of variability but generally inferior to that of NSE (FL).

As for the accuracy of the model performance (**Fig. 4.3b**), all precipitation products (except NARR) consistently generated a negative PBIAS (FL) for the Battle River basin (stations 29-32) in all seasons. This result suggests that in this basin precipitation errors play a dominant role in affecting the model accuracy in streamflow simulation. Although there are no adjusted precipitation gauges in the Battle River basin and therefore no direct assessment could be done, the above precipitation analysis showed that NARR generally overestimated the precipitation amounts while other datasets provided different degrees of underestimation (*Yassin et al.,* 2019b). The degree of accuracy of the streamflow simulation reflected that of the precipitation products. A similar positive association between the accuracies of precipitation products (*Yassin et al.,* 2019b) and their associated model performances was witnessed in the headwater of the Red Deer River basin where NARR generally showed a positive PBIAS (FL) (station 18) for the individual seasons (overall PBIAS (FL) of 21%).

Likewise, the general underestimation of precipitation amounts in Princeton and CaPA in four seasons (*Yassin et al.,* 2019b) directly propagated and affected the streamflow simulation in the headwater of the Upper South Saskatchewan River basin in which every PBIAS (FL) was negative (-67% and -35% overall average underestimation for Princeton and CaPA, respectively). In particular, the general overestimation and underestimation of streamflow simulation at station 5 followed the overestimation and underestimation of precipitation amounts by different

precipitation products (*Yassin et al.,* 2019b) for the individual seasons. However, such seasonally positive association was not strong in other parts of the headwaters, where the overestimation of precipitation amounts by CRU and GPCC in winter (station I (*Yassin et al.*, 2019b)) was followed by the underestimation of streamflow simulation in winter and overestimation in spring (stations 2 and 3). This could possibly be due to the lag time of the hydrological response to precipitation forcing. Furthermore, this positive association was dampened when going downstream and did not hold in other subbasins.

Most of the streamflow simulation was overestimated in the Upper North Saskatchewan River basin (station 26) in spring and winter, despite the underestimation of precipitation amounts in the datasets (*Yassin et al.,* 2019b). This overestimation is similar to that of the Lower North Saskatchewan River basin. This contradiction could be due to (1) the error propagated from the mountains in the Upper North Saskatchewan where a large portion of the flow is generated, and (2) the mixing effects of error from the precipitation products and other model errors. Additionally, the streamflow simulation was consistently underestimated in all seasons in the Prairie watershed of the Lower South Saskatchewan River basin (station 34), regardless of the accuracies of the precipitation datasets (*Yassin et al.,* 2019b). In this case, the errors in the precipitation products played only a small in affecting the streamflow simulation.

The aim of the experimental design here was to reveal the quality of the precipitation products from a hydrological perspective while trying to isolate the effect of any compensations from parameter uncertainty. The basic assumptions were as follows: 1) given the same model structure and process representation across the basin with one set of a priori parameter values, any differences in simulations shown in the above analysis would mainly come from precipitation errors, and 2) assuming the model structure and processes were represented correctly and the streamflow was measured with minimal uncertainty, any overestimations and underestimations of the precipitation amounts should symmetrically transfer to the errors in streamflow simulations. We observed that the model performance was not always in harmony with the errors assessed in the precipitation products e.g., an overestimation of streamflow while precipitation amounts from

different precipitation datasets were generally underestimated. Such cases, could imply that the errors from the precipitation products were outweighed by other errors. For instance, the results from the above streamflow-based precipitation comparison could have be affected by the choice of the a priori parameter set because the chosen parameter set might not have represented the correct values for the processes to function properly. Additionally, the baseflow representation in MESH with a conceptual bucket below the soil profile plays a major role in how well the low flows are simulated, directly affecting how well the low-flow-season performance metrics perform. Consequently, errors from the process representation were introduced into the results. However, we acknowledged the difficulty of segregating the effects of errors from different sources, given the complexities of each sub-basin across the SaskRB and the insufficient understanding of all the hydrological processes and human activities in the basin.

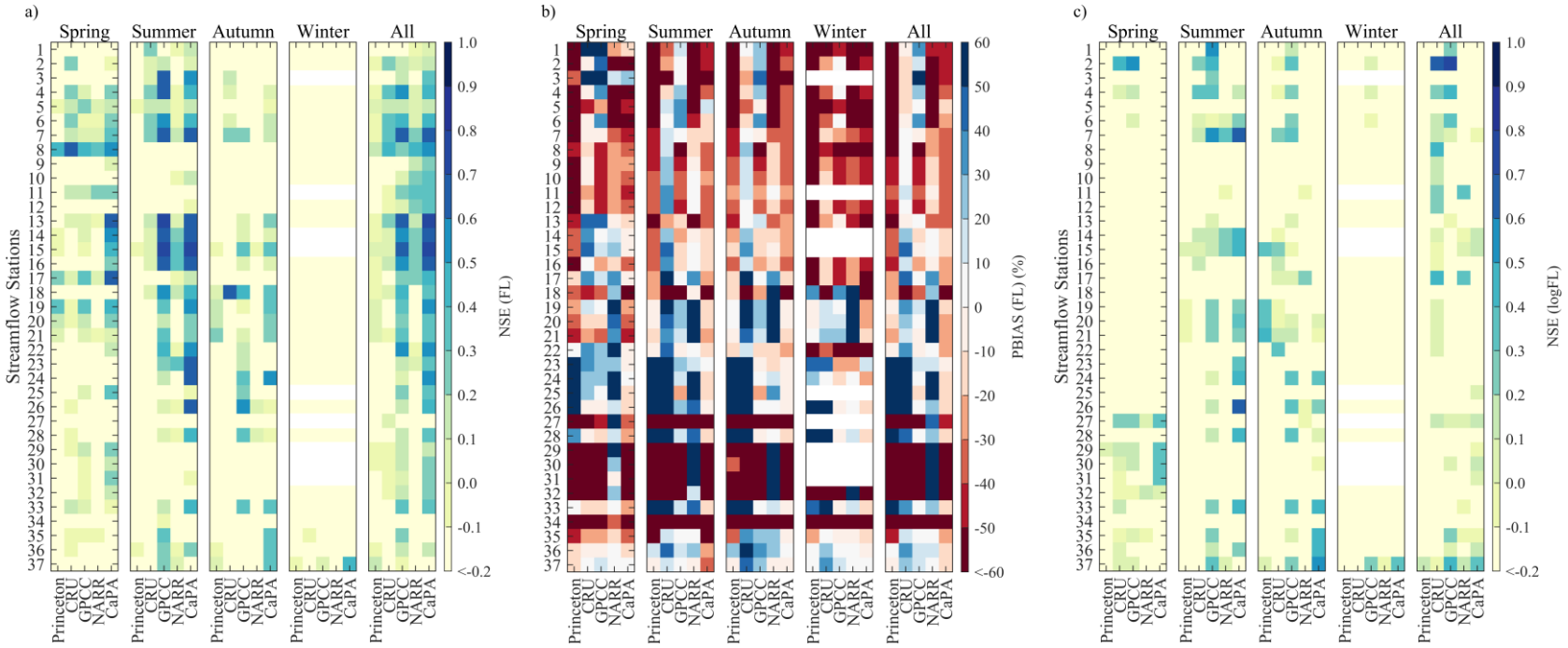


Figure 4. 3 Streamflow daily simulation performance metrics for different precipitation data (10 years streamflow simulation against observations of multiple stations) a) NSE (Flow) b) PBIAS (Flow), c) NSE (logFL). White indicates that no streamflow data are available as they have only seasonal observations. Stations 3, 11, 14, 15, 25, 27, 30, and 31 only have seasonal observations with no observations during winter

4.6.2 MESH model calibration under the best performing precipitation data

In the previous section's analysis, CaPA was identified as the best-performing precipitation dataset, and therefore, CaPA is used in this section for model calibration. As mentioned in section 4.5.3, a sub-set of parameters from MESH horizontal, vertical and routing component were selected for model calibration for two reasons: (1) to reduce the number of model evaluations needed in calibration, thereby reducing the computational cost, and (2) to avoid compensation between error sources and to avoid improving streamflow simulation at the cost of degrading other model output. For example, we clearly presented that CaPA underestimates precipitation, and we are wary of compensating for this underestimation by reducing the evapotranspiration amounts.

The parameter calibration results are presented in **Fig. 4.4** using NSE (FL), NSE (logFL), and PBIAS (FL) performance metrics. Each plot in **Fig. 4.4** has six components: (1 and 2) performance with a priori parameters during calibration and temporal validation periods (i.e., pre-cal and pre-val), (3 and 4) performance with calibrated parameters during calibration and validation periods (i.e., post-cal and post-val), (5 and 6) validation using independent stations (spatial validation) during calibration and validation periods (**Fig. 4.4**, these stations are indicated in italics and blue colors).

In the Upper South Saskatchewan River basin, model performance improved after calibration in all of the cases; the calibrated model even performed better in the validation period than in the calibration period. A possible reason for this is that the basin becomes wetter in the validation period. When greater precipitation amounts drive the model, it can more easily match the streamflow. This can be demonstrated by the fact that the two "wet" precipitation products (CRU and GPCC) were able to produce better streamflow simulations (especially in summer) than the "dry" CaPA without calibration (**Fig. 4.4**). The major improvement in model performance in the validation period is coming from spring, possibly because of reduced underestimation of CaPA and/or increasing temperatures that leading to more snowmelt and runoff.

A similar model performance was seen in the headwaters of the Bow River basin (stations 8 - 12), where the model produced better NSE (FL) during validation than in calibration. The improvement in model performance occurred in spring and summer in stations 10, 11, and 12, and in autumn and winter at station 8. MESH performed consistently well in the lower part of the Bow River basin where major irrigation areas, diversions, and upstream reservoirs are found. This shows that our newly developed modules of irrigation, reservoirs, and flow diversion were capable of capturing the regulated streamflow in this section of the basin. The model also performed well in the validation stations (stations 14 and 15) during both calibration and validation, implying that the global calibrated parameter sets were able to capture the hydrological dynamics of the basin in both space and time.

A different model performance was observed in the Red Deer River basin. The model performed well at both calibration and validation stations during the calibration period. However, the model failed to simulate the streamflow during the validation period (stations 18-21), with NSE (FL) ranging from 0.13 to negative values and with a high positive PBIAS (FL) of 57%. Furthermore, the model performance was worse after calibration in the validation period in which the model produced more streamflow than that observed in summer and autumn (**Fig. 4.4c**). This worsening model performance could be due to two factors: (1) a possible failure in transferring the global calibrated parameter sets over space and time; (2) the introduction of errors during model calibration when the model failed to properly represent some of the hydrological processes (e.g., partitioning of underestimated precipitation amounts with low evapotranspiration, leading to high streamflow). A similar model performance was observed in the Eastern Saskatchewan River basin (stations 36 and 37), where the calibrated results in the validation period were worse than uncalibrated ones. The reasons could be similar to those affecting the Red Deer River basin.

Although the model performed similarly poorly in the Battle River basin, the reasons for this were different. In this case, the underestimation of precipitation amounts played a major role, as discussed in the previous section. The model was able to reduce the negative PBIAS at the calibration stations after calibration but failed at the validation stations. Again, the hydrological

dynamics of this basin were not well captured by the transfer of global calibrated parameter sets. However, the calibrated streamflow simulations of these stations are generally of good quality and have good correlations with observations (**Table 4.7**).

Another model failure could be seen in the upper part of the Lower South Saskatchewan River basin (station 34) where the model performance was better without calibration in both the calibration and validation periods. Given the overestimation of precipitation amounts throughout the seasons (except summer), the model produced a high negative PBIAS (FL) for the individual seasons after calibration. As with the other case, this overestimation could be attributed to the improper representation of some processes in this part of the basin.

The above evaluation strategy enabled us both to reveal the ability of the model to capture the hydrological response across the basin and to assess the global multiple-station calibration method in transferring the parameter sets in space and time. Overall, the model was able to perform well across the SaskRB with significant improvement in the median of NSE (FL) and NSE (logFL) in both the calibration and validation periods, compared with those obtained pre-calibration. Similarly, the median of the PBIAS (FL) of all stations was reduced by more than 5 % in all cases. Additionally, the globally calibrated parameter set was able to provide reasonable streamflow simulations over validation stations and during the validation period. Despite some failures, the global parameterization generally achieved better model performance across the basin in which nearly 60% and 30% of the stations resulted in NSE (FL) and NSE (logFL) > 0.5, respectively.

Despite these encouraging results, unsurprisingly, there were regional differences in model performance where the model failed to capture the hydrological regimes of some subbasins. Given the considerable dry bias in CaPA, it was not expected that the model could match the observed streamflow and result in a high negative PBIAS (FL). This was found to be especially true in subbasins where precipitation errors played a dominant role (e.g., Battle). Poor model performance during the validation period in some subbasins (e.g., Red Deer) might have occurred because the model was calibrated mainly to data measured dry (drought) years (2003 – 2005) in the Prairies.

Other precipitation and/or hydrological regimes were, therefore, not able to be captured by calibration.

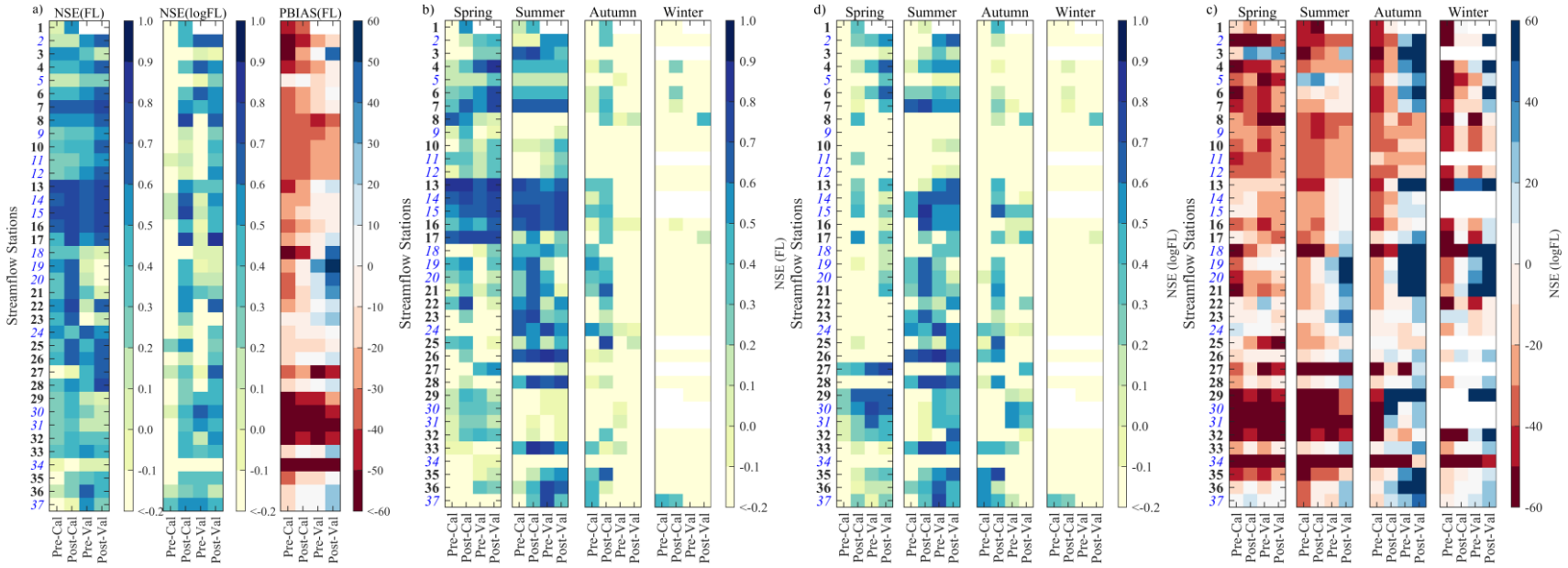


Figure 4. 4 Streamflow daily simulation performance metrics before and after calibration. White indicates that no streamflow data are available as they have only seasonal observations. Calibration station numbers has bold font weight and spatial validation station numbers are italicized and in blue color. Stations 3, 11, 14, 15, 25, 27, 30, and 31 only have seasonal observations with no observations during winter.

4.6.3 Model validation on other fluxes and stores

This section describes further validation of the model performance conducted on other hydrologic fluxes and stores using additional observations of evapotranspiration (ET) and the total water storage (TWS) anomaly. Measurements of evapotranspiration are scarce and limited to flux tower observations. Modeled ET is often compared to flux tower observations with a very limited spatial coverage and/or to satellite-based gridded ET estimates. In this study, consistency in MESH simulated ET was assessed using NDVI-based ET estimates and point ET estimate in two-flux towers. The simulated TWS was compared against the GRACE satellite TWS observations. To reveal the level of performance differences, the validation of ET and TWS is presented for both before and after model calibration.

Fig. 4.5 presents a box-and-whisker plot to compare daily observed and simulated ET for the period 2003-2009 at two flux tower sites: Old Jack Pine (OJP) and Old Black Spruce (OBS). Both sites are located in the downstream Boreal Plain (**Fig. 4.1c**). The flux-tower observations were compared to a needleleaf forest tile near the flux-tower sites. The majority of ET at both sites occurs from April to October, with a large portion in summer (June, July, and August). The median ET is close to zero for the low-temperature months from November to March. **Fig. 4.5a** shows that the pre and post-calibrated median and interquartile ranges of ET followed the observed seasonal ET pattern well. However, overestimation by MESH was observed from May to July, and the interquartile ranges were larger than those of the observed ranges. The median of post-calibrated ET was closer to observation than pre-calibrated ET.

Fig. 4.5b shows modeled and observed ET comparisons for OBS, with both pre- and post-calibration results, indicating comparable ET estimations against observations for median and interquartile ranges. Unlike OJP, ET underestimation was observed for the OBS site for July, August, and September. The difference between simulated and observed ET for the flux-tower sites is in part associated with scale mismatch between the MESH tile-scale simulations and flux point measurements (typically the coverage of a flux tower is at a scale of 100 m), bias in

meteorological forcing, and process representation of soil and vegetation in the model. Additionally, the observation from flux towers involves some adjustment for energy balance closure (*Barr et al.,* 2012). Nevertheless, the overall performance at both flux towers showed similar performance quality as in a previous study by *Davison et al.* (2017) which conducted detailed MESH model calibration for a watershed where both observations are located. The results of our study indicates good performance of our globally optimized parameters values over the flux towers' region.

Regarding the gridded comparison of ET and TWS, a simple comparison was made by sampling the values of GRACE TWS and NDVI-based ET at the MESH grid scale using the nearest grid point approach. **Fig. 4.6a** and **Fig. 4.6b** present gridded correlation coefficients between NDVI-based monthly ET and the MESH-simulated monthly ET before and after calibration, respectively. The comparison showed a reasonably good agreement between simulated ET and NDVI-based ET for both before and after calibration. Both had a high positive correlation (> 0.8) over a large portion of the basin. The patterns of low correlation were concentrated around the same regions in the basin, the majority of which were in the lower portion of the Upper South Saskatchewan, with a small portion of the lower end of the Bow and upstream of the Lower South Saskatchewan subbasins. The low correlation region is potentially related to irrigated areas (**Fig. 4.1d**) where more water is available from irrigation water diversion. However, other factors could also contribute to the difference between NDVI-based ET and MESH ET, such as variation in the ET estimation methods and scale mismatch.

The gridded correlations between GRACE TWS and MESH-TWS (**Fig. 4.7a** and **Fig. 4.7b**) showed a wide range of performance, but the general patterns of pre- and post-calibration performances are in agreement. A very good agreement was also observed between the basin averaged time series of the simulated and observed TWS anomaly (**Fig. 4.7c**). The calibrated MESH gridded TWS correlation results showed larger coverage (more grid cells) of high correlation than pre-calibration TWS correlation result. The high correlation values were observed around the central part of the SaskRB, including the Lower North and South Saskatchewan

subbasins as well as the lower portion of the Battle and Central North Saskatchewan subbasins. The headwater subbasins (Upper South Saskatchewan, Bow, and Red Deer) showed high correlation around the upstream and downstream ends of the subbasins. However, the gridded results revealed that the model did not perform well in some places, particularly in the Eastern part of the SaskRB, possibly because the process representation over open waters (lakes) is deficient in MESH, as the inferior correlations are scattered around the water bodies of the SaskRB. Besides the MESH deficiency, it is notable that there are varying levels of uncertainty that come with GRACE data related to issues of scaling, filtering, and the removal of ocean, atmosphere and isostatic rebound signals (*Seo et al.*, 2006; *Landerer and Swenson*, 2012).

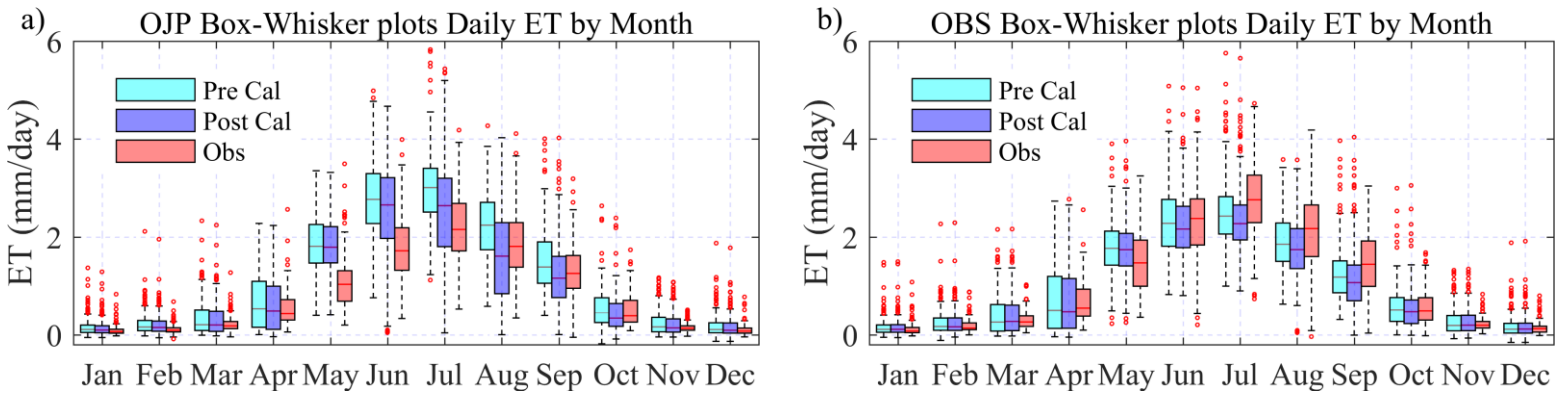


Figure 4. 5 Box plots of observed and simulated pre and post calibrated daily ET in mm day^{-1} for the period of 2003-2009 (a) Old Jack Pine site, (b) Old Black Spruce site.

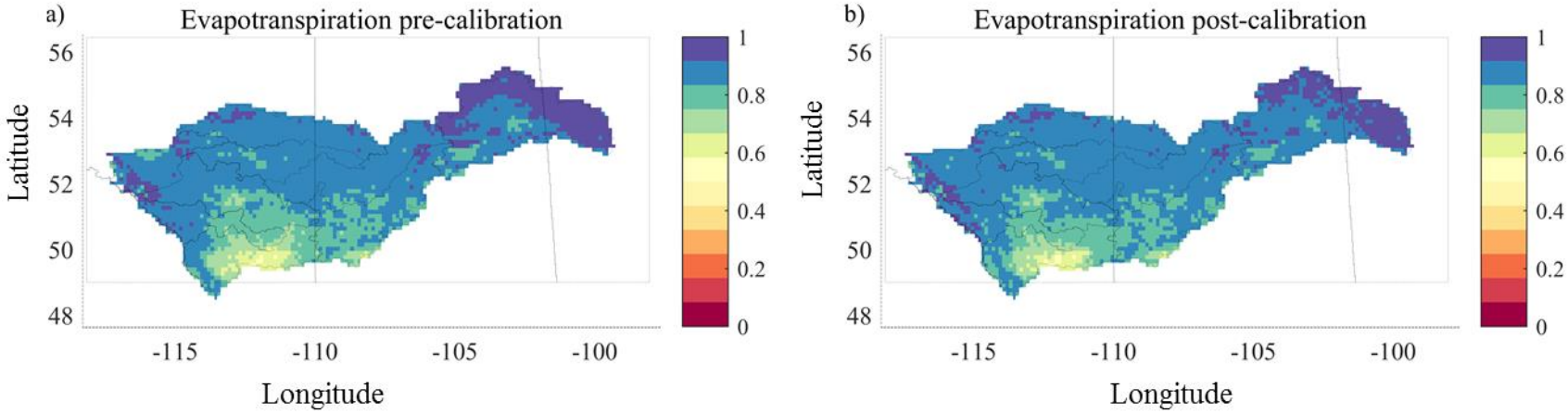


Figure 4. 6 MESH evapotranspiration monthly simulations (2003 – 2010) performance (a) correlation of ET pre-calibration against NDVI-based ET, (b) correlation of ET post-calibration against NDVI-based ET. The *x*-axis shows longitude and *y*-axis shows latitude.

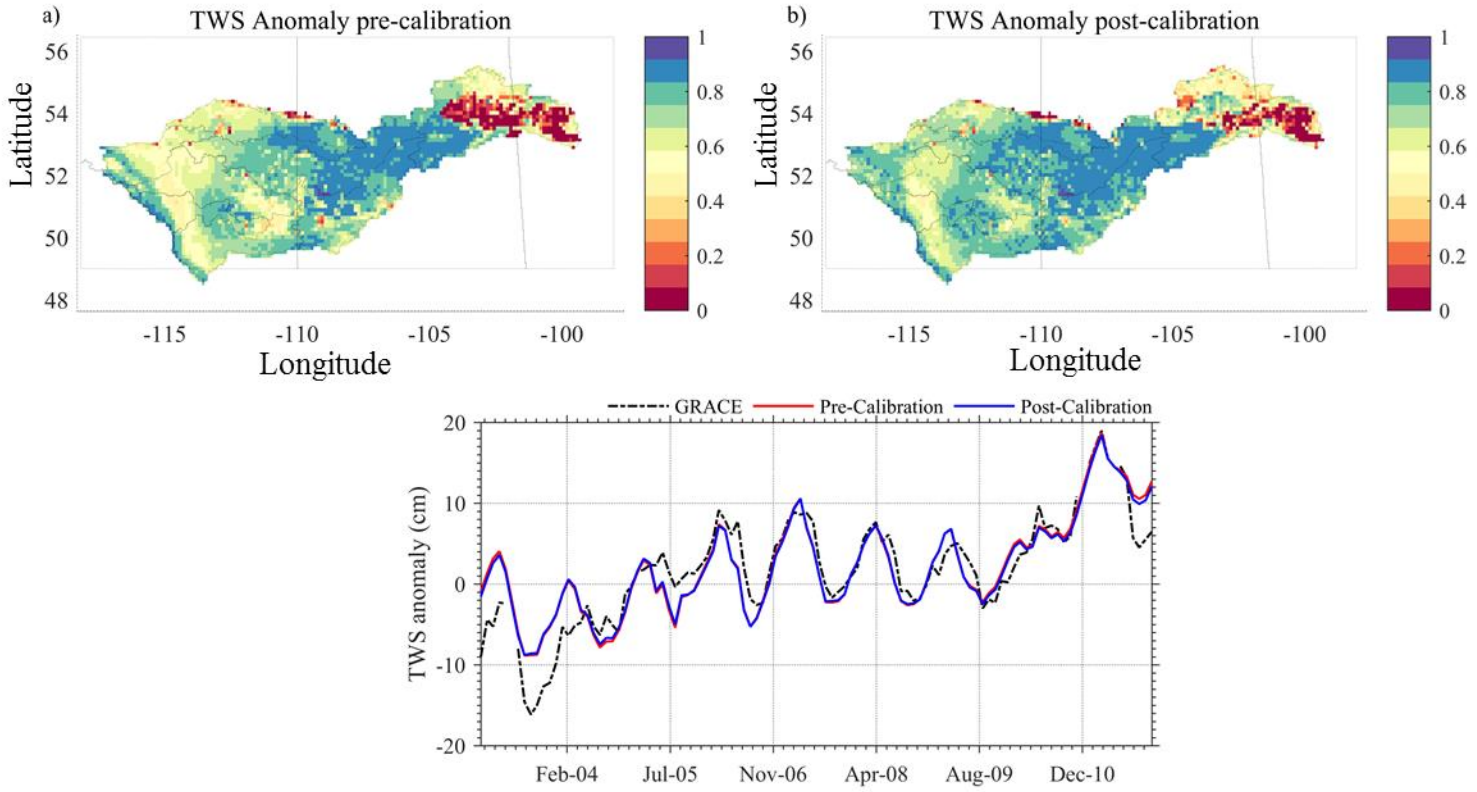


Figure 4. 7 MESH model performance (a) correlation of TWS pre-calibration (b) correlation of TWS post-calibration (c) Basin average simulated TWS (cm) comparison for pre-calibration and post-calibration along with GRACE-based monthly TWS.

4.7 Conclusions and implications

In this work, we tested the capability of the MESH Hydrologic-Land Surface Model in capturing the hydrological dynamics of the SaskRB using a multi-criterion, multi-station calibration approach, recognizing the input uncertainties from the climate forcing data and the complexities of the river system. To minimize the effects of error compensation from model structural and parameter uncertainties and to reduce computational cost, we used the different precipitation datasets to drive the MESH calibration with a default parametrization to evaluate their performance in reproducing streamflow (section 4.6.1). CaPA was the overall best-performing precipitation product, so we used it to calibrate MESH against observed streamflow. We evaluated the model by using four sets of criteria to test the transferability of the global calibrated parameter sets in space and time (section 4.6.2). Finally, we took a further step to evaluate the model's ability in reproducing other water budget components by comparing them with additional information (section 4.6.3). We conclude that:

1) The streamflow simulation based on multiple precipitation product showed the quality of precipitation products had a direct and immediate impact on the headwaters of the basin but the effects were dampened when moving downstream. However, such associations did not hold in some subbasins, reflecting the possibility that other errors (e.g., model structure and process representation) had potentially outweighed or offset the errors from the precipitation products.

2) In general, MESH was able to capture the hydrological response across the basin using the global multiple-station calibration method. Despite some failures, the global parameterization generally achieved better model performance across the basin in which ~ 60% and 30% of the stations resulted in NSE (FL) and NSE (logFL) > 0.5, respectively. Given the considerable dry bias in CaPA and the complexity of SaskRB, it is not surprising that there were regional differences in model performance, where the model failed to capture the hydrological regimes of some subbasins.

3) Further validation with complementary water budget component products (GRACE, NDVI-based ET, and two flux tower sites) showed good representation of ET for flux tower observation and good correlation with NDVI-based ET and GRACE observations. However, the validation revealed MESH's potential deficiency in capturing water storage over open water areas (i.e., in the Eastern Saskatchewan River basin).

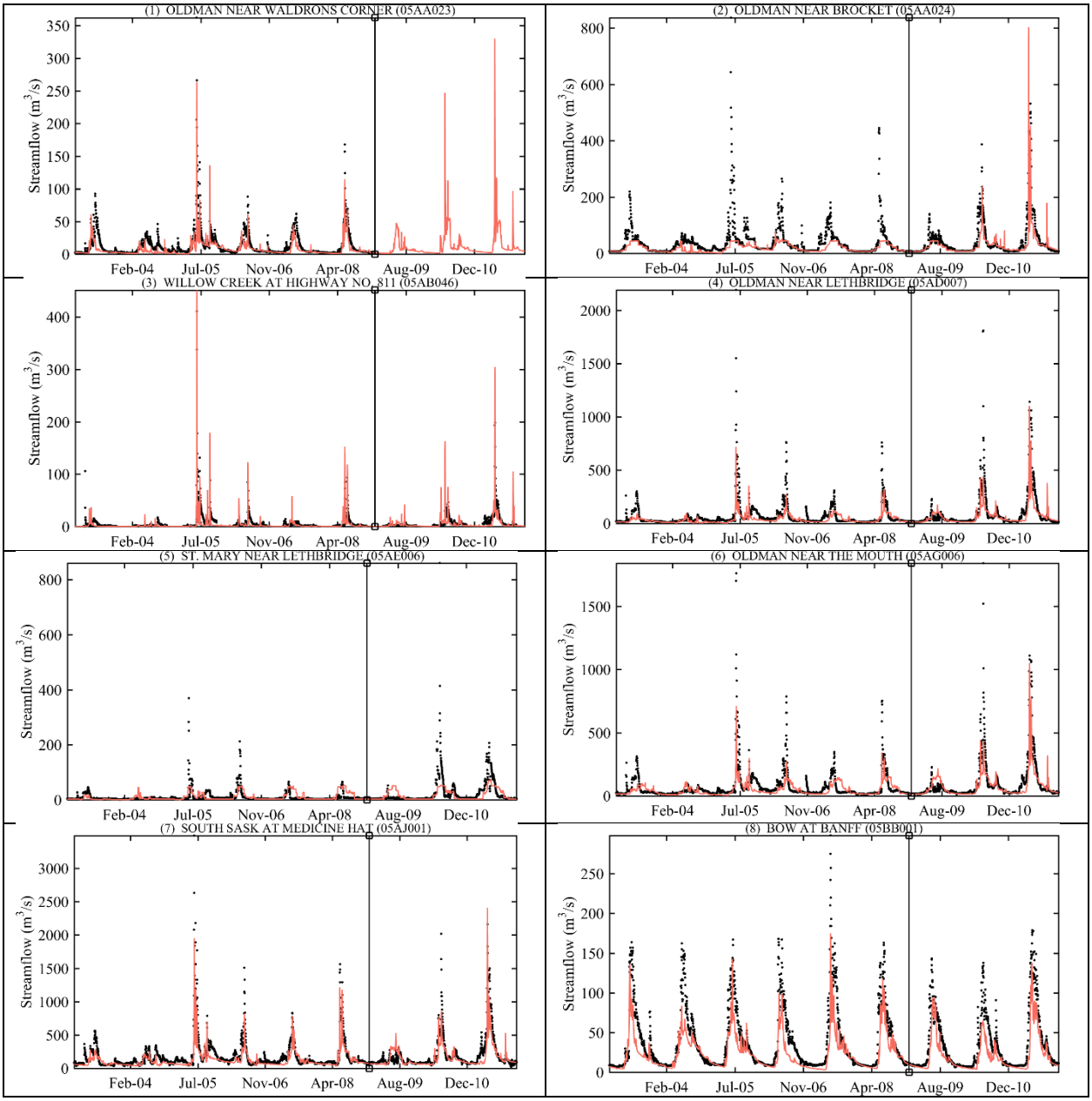
The broader implications of this work include the following:

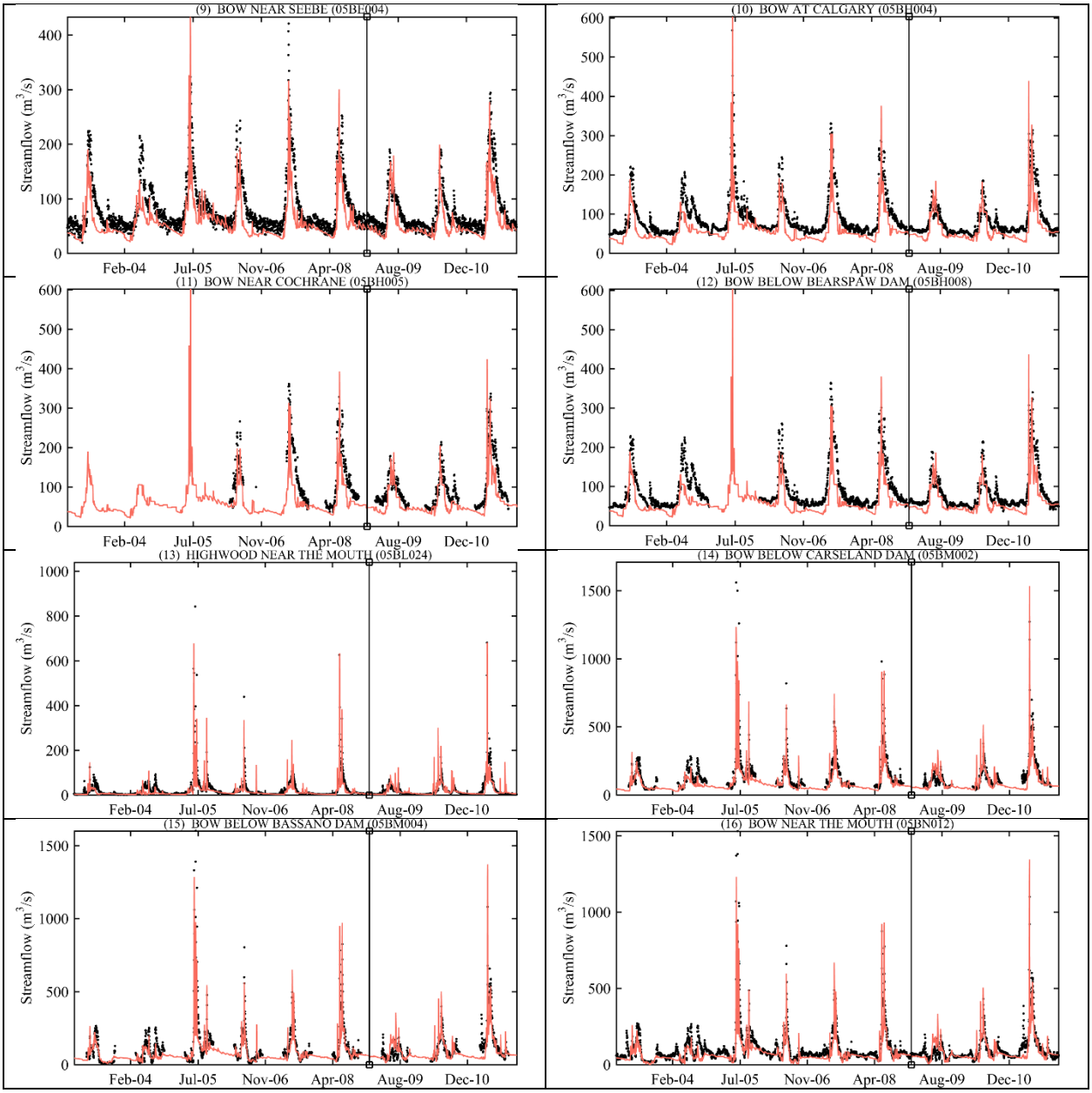
1. Generally speaking, diagnosing model failures and potential error compensations among different model components in the presence of precipitation uncertainty is more challenging in the case of conceptual global and catchment-scale hydrological models, in which different processes are approximated conceptually in a simplified way and parameter calibration is key to reliable streamflow simulation (*Schmeid et al.*, 2014). On the contrary, diagnosing limitations of process-based H-LSMs faces less risk of error compensation and can be possibly done without calibration, because most H-LSMs parameters have a direct physical interpretation, which introduces the possibility to define realistic default parametrization. This study showed that the approach of default parameterization could be used effectively to evaluate and identify suitable climate forcing and to reduce the risk of error compensation between different model components in the course of calibration. This reveals the potential benefits of process-based H-LSMs over conceptual hydrological models in the identification of the sources of errors and future directions for model improvement.
2. Similar to hydrological models, calibration can help us achieve a better performance of H-LSMs, because of the inconsistency of scales between measurements and modelling grid, simplification of the surface heterogeneity representation and hydrological processes, parameter uncertainty in acceptable interval values, etc. For instance, *Nasonova et al.* (2009) demonstrated calibration significantly improved the H-LSM streamflow simulation for several Model Parameter Estimation Experiment (MOPEX) catchments, and the performance of H-LSM on streamflow simulation after calibration was closely matched

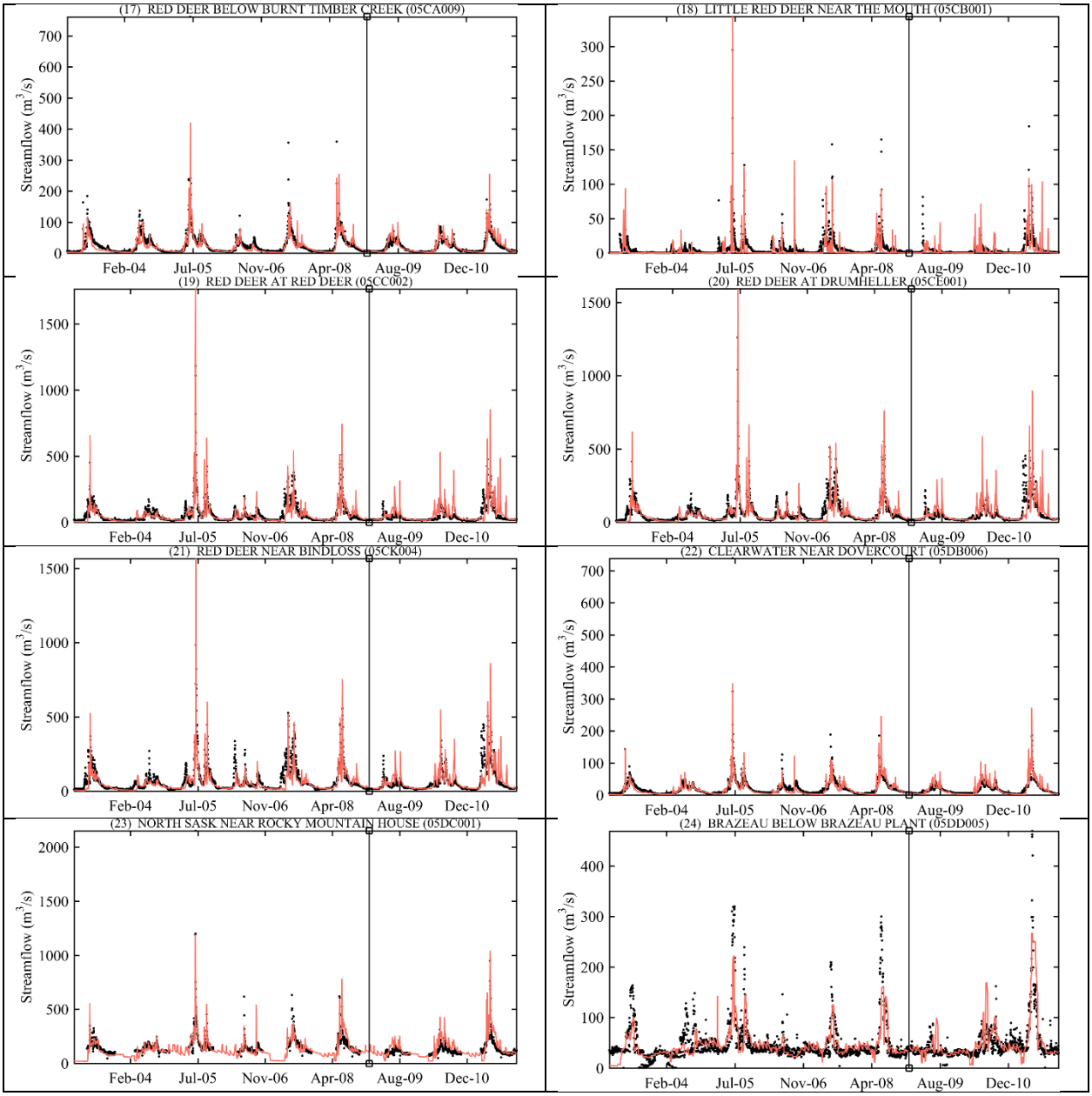
the hydrological models used in *Duan et al.* (2006). In a similar context, this study affirms that the utilization of effective calibration (parametrization) in an H-LSM makes it a robust modelling system, capable of producing improved streamflow simulation for large-scale basins with complex hydrological processes and highly managed environment. The H-LSM parameterizations enable the possibility of modelling ungauged watersheds through transferring optimal parameters, for example based on similarity of land cover. However, care must be taken with the number of parameters to be calibrated, the usage of observations used in the calibration/validation processes, and the strategy of the calibration method.

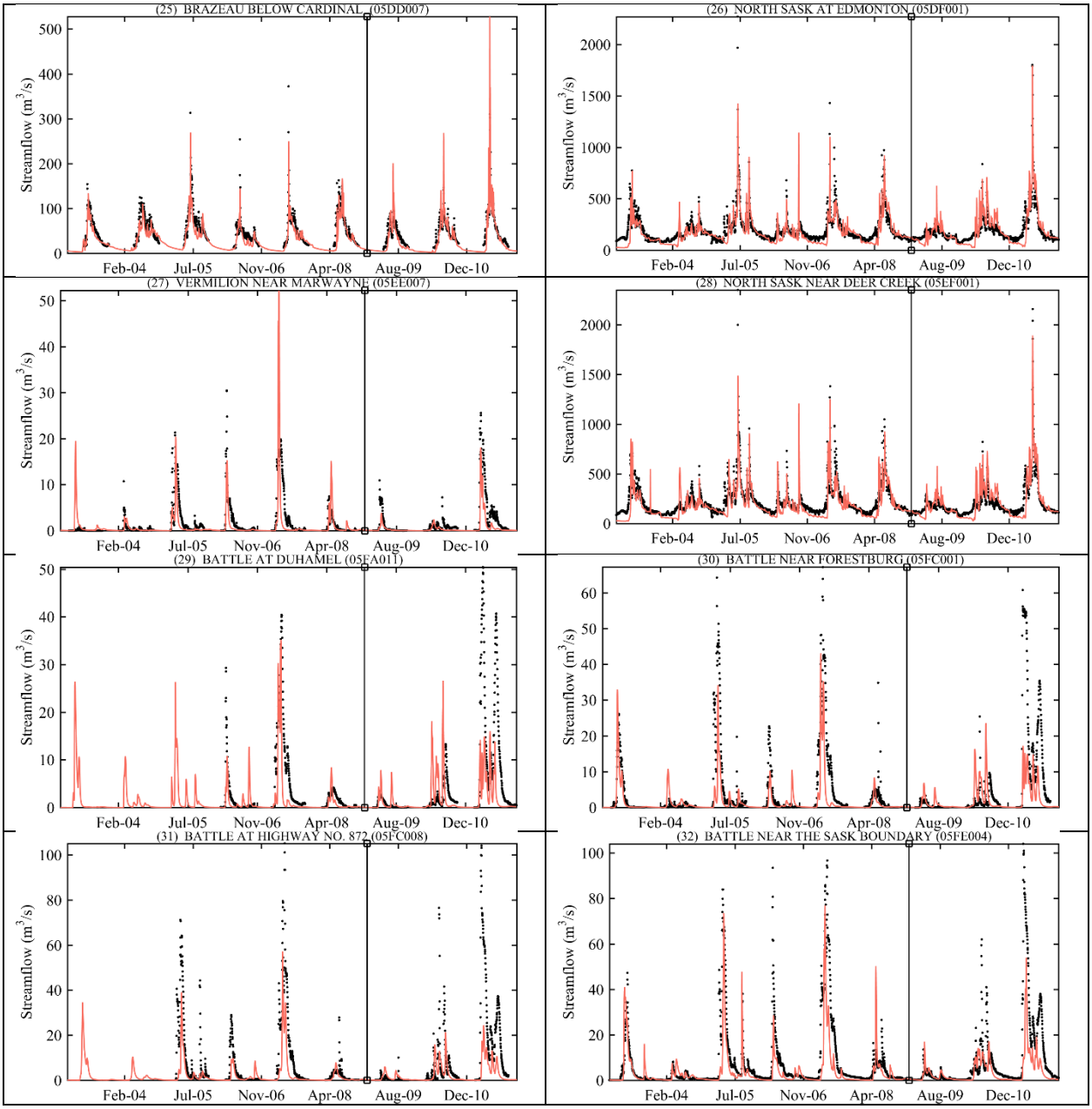
3. Given the high-dimensionality of parameter spaces and the lack of confidence in defining the parameter values, one might argue that more (or even all) parameters should be calibrated to obtain better model performance. While model parameters are usually calibrated to streamflow only, calibrating more parameters could increase the risk of over-fitting and the chance of introducing errors into other model stores and fluxes. Therefore, except the case of experimental watersheds with comprehensive observations, calibrating many parameters should not be encouraged as this essentially reduces the benefits of the 'process-based' nature of H-LSMs. With the increasing availability of observational information on hydrological components such as GRACE total water storage anomaly and satellite-based ET, one might utilize these data during data assimilation processes or calibration alongside with observed streamflow. Uncertainties are often inherent within these data, but rigorous assessments of the error distribution of these data are limited. This is especially true in the SaskRB (and in Canada generally), where ground observations are very sparse or simply non-existent. Thus, our study only utilized these data for temporal and spatial validation. Nonetheless, a more concerted effort is needed to assess and quantify the error characteristics of these data such that they can be better utilized in H-LSM development and testing.

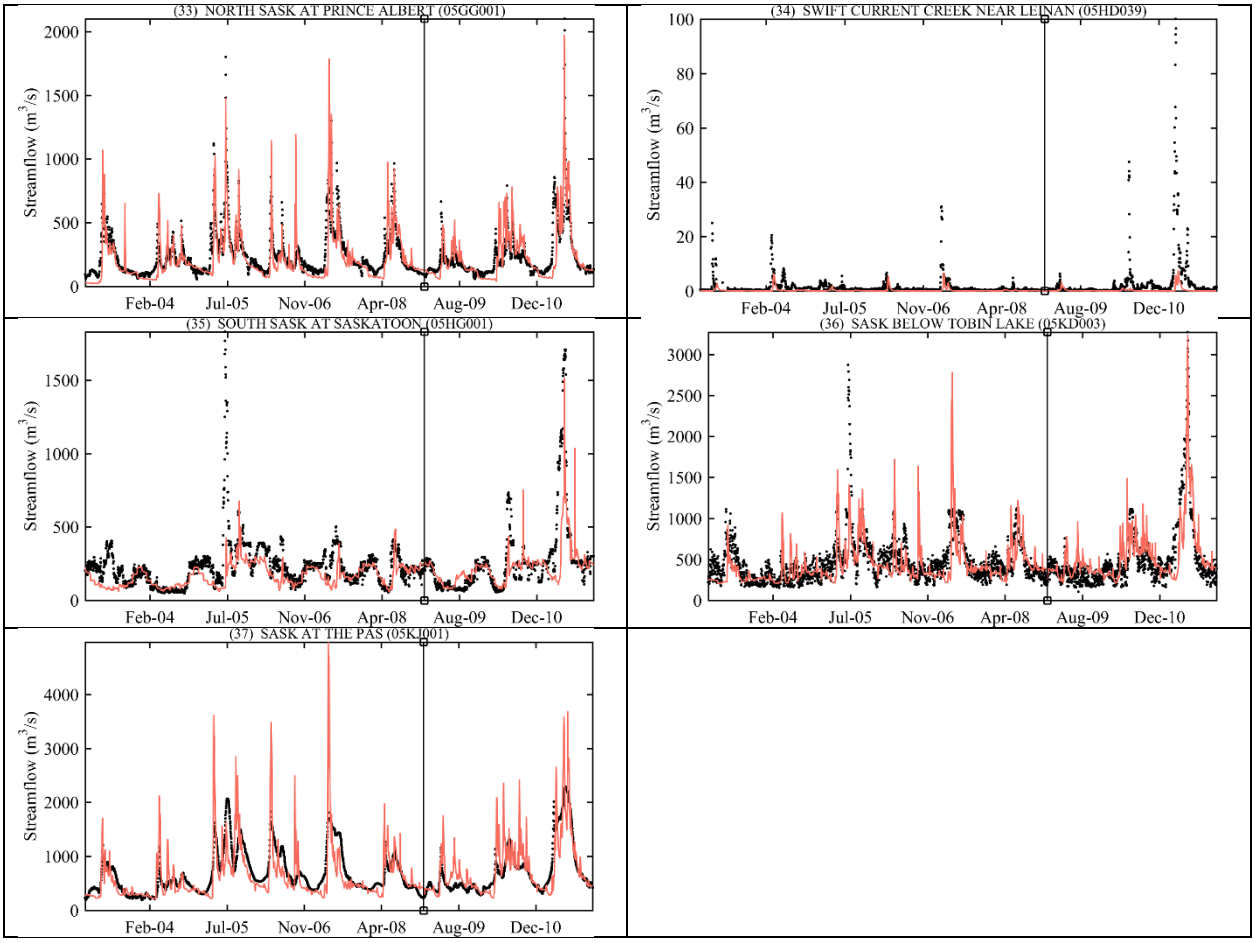
Table 4. 7 Simulated (light-red) and observed (black) streamflow of selected stations in SaskRB. The calibration and validation period is separated by vertical line at the end of 2008.




(9) BOW NEAR SEEBE (05BE004)
(10) BOW AT CALGARY (05BH004)
(11) BOW NEAR COCHRANE (05BH005)
(12) BOW BELOW BEARSPAW DAM (05BH008)
(13) HIGHWOOD NEAR THE MOUTH (05BL024)
(14) BOW BELOW CARSELAND DAM (05BM002)
(15) BOW BELOW BASSANO DAM (05BM004)
(16) BOW NEAR THE MOUTH (05BN012)
Streamflow (m3/s)
Feb-04
Jul-05
Nov-06
Apr-08
Aug-09
Dec-10


(17) RED DEER BELOW BURNT TIMBER CREEK (05CA009)
(18) LITTLE RED DEER NEAR THE MOUTH (05CB001)
(19) RED DEER AT RED DEER (05CC002)
(20) RED DEER AT DRUMHELLER (05CE001)
(21) RED DEER NEAR BINDLOSS (05CK004)
(22) CLEARWATER NEAR DOVERCOURT (05DB006)
(23) NORTH SASK NEAR ROCKY MOUNTAIN HOUSE (05DC001)
(24) BRAZEAU BELOW BRAZEAU PLANT (05DD005)
Streamflow (m3/s)
Feb-04
Jul-05
Nov-06
Apr-08
Aug-09
Dec-10


(25) BRAZEAU BELOW CARDINAL (05DD007)
(26) NORTH SASK AT EDMONTON (05DF001)
(27) VERMILION NEAR MARWAYNE (05EE007)
(28) NORTH SASK NEAR DEER CREEK (05EF001)
(29) BATTLE AT DUHAMEL (05EA011)
(30) BATTLE NEAR FORESTBURG (05FC001)
(31) BATTLE AT HIGHWAY NO. 872 (05FC008)
(32) BATTLE NEAR THE SASK BOUNDARY (05FE004)
Streamflow (m3/s)
Feb-04
Jul-05
Nov-06
Apr-08
Aug-09
Dec-10


(33) NORTH SASK AT PRINCE ALBERT (05GG001)
(34) SWIFT CURRENT CREEK NEAR LEINAN (05HD039)
(35) SOUTH SASK AT SASKATOON (05HG001)
(36) SASK BELOW TOBIN LAKE (05KD003)
(37) SASK AT THE PAS (05KJ001)
Streamflow (m3/s)
Feb-04
Jul-05
Nov-06
Apr-08
Aug-09
Dec-10

CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Summary and conclusions

In this dissertation, contributions to methodologies (approaches) dealing with some of the challenges in developing the enhanced capability of Hydrologic-Land Surface Model (H-LSM) was presented. The developments in H-LSMs aim to contribute improved simulations of model stores and fluxes that are essential to manage water resources, to examine the impact of future environmental changes, and to advance forecasting skills for hydrological extremes. The proposed methodology, and the findings in this dissertation are of global significance and are applicable to any region and H-LSM. The summary and conclusions obtained from each component addressing the research questions stated in section 1.3 are outlined below.

- Chapter 2 presented the issues on parameter equifinality and possibly false model state response that arises from calibrating complex models (such as H-LSMs) only to streamflow observations. This was demonstrated using the multi-criteria global sensitivity analysis and an optimization approach to compare streamflow-based calibration versus multivariate calibration (based on streamflow and GRACE Total Water Storage (TWS) anomaly). The sensitivity analysis of model parameters revealed that the parameters that are the dominant controls of water storage across the basin (e.g. soil properties of deeper soil layers) may be less influential on streamflows. The multi-criteria calibration with and without GRACE data showed that GRACE introduces a new angle to the evaluation of model performance, indicating the weakness of streamflow-based error metrics in adequately identifying model parameters. The added new dimension by GRACE contributes to addressing the issue of non-uniqueness of the optimal parameter values, to constraining the behavioral parameter space, and subsequently, to guiding the dynamical behavior of the model's over time. Overall, this study highlighted the need to better constrain the model's degrees of freedom by looking at the model internal state trajectories in addition to streamflow.

- Chapter 3 provided a thorough review of existing reservoir operation models and their limitations within Catchment Models (CMs), Global Water Security Models (GWSM), and H-LSMs. Then, the chapter presented an improved reservoir operation model (reducing existing model limitations), called Dynamically Zoned Target Release (DZTR), developed using parametric piecewise linear relationships between reservoir storage, inflow and release to mimic actual reservoir operations. The developed model allows two types of parameterization, one based on the distributions of historical storages and releases to generate the so-called "generalized parameters" and the other one based on direct calibration to observed storage and release time series via multi-objective optimization. The result with both parameterization approaches showed that the proposed reservoir model leads to improved reservoir storage and release simulation accuracy compared to the other widely used approaches. Further tests on integration of this reservoir model into MESH (Canadian H-LSM) improved the overall model performance compared to the traditional methods of accounting for reservoirs in H-LSMs. Such an integration provides a successful example for improving the representation of reservoir operation in CMs, LSMs and GWSMs.
- Chapter 4 examined the benefit of calibrating H-LSM parameters, and H-LSM capability to simulate a cold-region, large-scale basin (the 406,000 km^2 Saskatchewan River Basin (SaskRB)) with complex hydrological processes and extensive water management regulation. The tests were conducted based on a Canadian H-LSM called MESH and driven by better quality precipitation data that were selected by evaluating multiple reanalysis climate forcing products. The success of MESH was evaluated by comprehensive criteria defined using a multicriteria, multi-station calibration approach, and validation based on the ability in reproducing other water budget components by comparing against evapotranspiration flux data and satellite based storage anomaly and evapotranspiration data. The benefits of calibrating MESH parameters was benchmarked by comparing with a MESH simulation based on default parameterizations. The results revealed: (1) reanalysis

precipitation data contain considerable errors, and selecting a reliable precipitation product among many available reanalysis precipitation datasets is very important for good performance of MESH. (2) Despite general underestimation, the Canadian Precipitation Analysis (CaPA) product performed well in minimizing precipitation error characteristics and simulating streamflow across SaskRB. (3) The global multi-station calibration on streamflow provided better MESH performance across SaskRB, and despite some failures in some regions, the calibration enhanced the default parameterization MESH streamflow simulation performance. (4) Finally, the validation on evapotranspiration and water storage anomaly showed good representation of spatiotemporal evapotranspiration and storage anomaly outputs. However, the validation revealed potential limitations in representing water storage over lake areas.

Overall, the demonstrated capability of an H-LSM in modelling a complex, highly-managed large-scale basin indicates its potential to examine impacts of future climate and land-cover changes, and impacts on water resources management. Being an embedded component of climate and numerical weather prediction models is one of the assets for H-LSMs to seamlessly evaluate climate change impacts and perform flood forecasting and drought monitoring. Moreover, the modular organization of H-LSMs enhances the flexibility of incorporating new model process components. For example, the implementation of reservoir operations and estimation of irrigation demand can be readily updated when new information is available. This facilitates the generation of different scenarios to evaluate the performance of existing reservoir operations under future changes and to assess the possibility of readjusting reservoir operation targets for adapting to future changes. Similarly, the generation of future scenarios to examine the effects of land-cover changes such as the expansion of irrigated areas, glacier retreats, deforestation, and forest fires is possible by altering the gridded model setting. Yet, it is still a challenge to dynamically represent the land-cover changes within the model. In addition, the ability of H-LSMs to output simulated variables regarding soil, vegetation, and snow at a sub-grid scale (tile-based) provides finer details for better

visualization over a region or a basin. As a result, H-LSMs are not only used as a modelling tool but also can serve as a platform to support policy making and water resources management.

5.2 Future work directions

The contributions in this dissertation have addressed only a sub-set of challenges posed in improving H-LSMs. Future work is suggested in the following areas based on this dissertation's perspective and other immediate needs to further improve H-LSMs.

- Future research may be directed towards the inclusion of a range of other remotely-sensed data of state and flux variables in a unified multi-criteria global sensitivity analysis and optimization framework for H-LSMs development and calibration at continental and global scale. In parallel with this suggestion, further research is needed to demonstrate the readiness of a range of remotely sensed data for the purpose of constraining H-LSM model parameters, this is because there might be inconsistency (incompatibility) and observational uncertainty issues between remotely sensed data and model outputs (*Wagner et al.,* 2007). For example, remotely sensed soil moisture data are limited to the shallow top soil layer, while the unsaturated zone can extend to several meters in depth. Another example, the GRACE native gravity anomaly observations passes through several steps to develop the water equivalent storage anomaly over land, creating a combination of many uncertainties and errors. (Based on Chapter 2)
- Many applications of global sensitivity analysis in H-LSM are applied to a single watershed with one streamflow station and few land-cover types (usually in H-LSM more land cover means more parameters). For better understanding of H-LSM parameters' roles for large-scale basins, it is important to evaluate parameter sensitivity based on distributed model output variables and distributed evaluation metrics (*Gupta and Razavi,* 2018). The distributed model output and metrics should include multiple streamflow stations and distributed output on complementary model output such as soil moisture and evapotranspiration. For a more comprehensive approach of determining model parameters

during snow dominated seasons and rain dominated seasons, the distributed sensitivity analysis experiments can be extended to account for the seasonal model output and evaluation metrics information (*Razavi and Gupta,* 2019). Additionally, the sensitivity analysis result for large number of H-LSM parameters need to be grouped into more informative classification (e.g. high, medium, and low sensitive parameters), for example this can be achieved using a strategy presented in the study of *Sheikholeslami et al.* (2019). (Based on Chapter 2)

- Reservoirs are one of the components that control the hydrologic cycle, and reservoir storage data are very important for reservoir operations and water allocation as demonstrated in Chapter 3. Hence, available reservoir storage (level) data are needed to be a part of model development and evaluation framework. For regions with limited reservoir storage data, further research work is needed to expand the availability by examining the utility of other data sources, such as those derived from satellite-based observations (*Savtchenko et al.,* 2004; *Garambois and Monnier,* 2015; *Gao et al.,* 2012) and using area-volume relationship approximated by regular geometric shapes (*Yigzaw et al.,* 2018). Progress on such types of reservoir storage dynamics estimation is becoming available as demonstrated in the recent study of *Busker et al.* (2019) which estimated storage data for 137 reservoirs across the globe. On a related note, further initiatives are needed to make reservoir operation rules as open public data (Alberta, Canada and Colorado, the USA are leading examples in this regard). The reservoir operation approximation though Dynamical Zoned Target Release (DZTR) model (Chapter 3) requires further testing on newly available dataset such as *Busker et al.* (2019). There is a need for an examination of a direct one- and/or two-way coupling of Water Management Models (WMMs) with CMs and LSMs. The coupling test of WMMs with CMs and LSMs will potentially improve development of a seamless coupled framework for the simulation of natural-engineered watershed systems. (Based on Chapter 3)

- Recommendations for future work regarding the Canadian H-LSM are presented as follows: (1) Seasonal model performance should become a common practice for model evaluation and calibration to emphasize model quality and limitation in representing snow- and rain-driven processes, separately. (2) Improving lake and wetland representations in CLASS. (3) Integrating groundwater module instead of using conceptual bucket baseflow for more realistic representation of storage dynamics in the basin and streamflow simulations. (4) Improving precipitation data (perhaps through bias corrections) for better diagnosing of model limitations. (5) Climate forcing, particularly precipitation reliability often evaluated against climate point observations, which limits the understanding of spatial structure or integrated quality of the products. Comparison studies should evaluate the reliability of forcing data directly against climate point observations and indirectly with the performance in streamflow simulation at basin, continental, and global scales. (6) Precipitation datasets from interpolation and reanalysis sources showed different qualities across spaces and time. To take full advantage of all datasets, a systematic assimilation of these datasets is needed. One direction in this regard could be creating a heuristic approach that weighs each source for each grid cell, and the weight can be estimated based on streamflow and other performance criteria. (7) The benefit of and strategies for, calibrating gridded soil texture need to be carefully revisited. For example, to retain the a priori spatial pattern of the observed soil texture, the parameterization can be done by perturbing gridded soil texture observations using land-cover units as a perturbation scale instead of grid by grid perturbation (each land-cover unit tends to have similar soil texture types and allows to retain the spatial pattern while perturbing the soil texture percentages). (Based on Chapter 4)

REFERENCES

Abbaspour KC, Rouholahnejad E, Vaghefi S, Srinivasan R, Yang H, Kløve B. 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology* 524: 733–752 DOI: 10.1016/j.jhydrol.2015.03.027

Abbott MB, Bathurst JC, Cunge JA, O'Connell PE, Rasmussen J. 1986. An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology* **87** (1–2): 61–77 DOI: 10.1016/0022-1694(86)90115-0

Abdulla FA, Lettenmaier DP, Wood EF, Smith JA. 1996. Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River Basin. *Journal of Geophysical Research: Atmospheres* **101** (D3): 7449–7459 DOI: 10.1029/95JD02416

Adam JC, Haddeland I, Su F, Lettenmaier DP. 2007. Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers. *Journal of Geophysical Research* **112** (D24): D24114 DOI: 10.1029/2007JD008525

Adam JC, Lettenmaier DP. 2003. Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research: Atmospheres* **108** (D9): 4257 DOI: 10.1029/2002JD002499

Alcamo J, Döll P, Henrichs T, Kaspar F, Lehner B, Rösch T, Siebert S. 2003. Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrological Sciences Journal* **48** (3): 317–337 DOI: 10.1623/hysj.48.3.317.45290

Andreadis KM, Clark EA, Lettenmaier DP, Alsdorf DE. 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters* **34** (10): L10403 DOI: 10.1029/2007GL029721

Archfield SA, Clark M, Arheimer B, Hay LE, McMillan H, Kiang JE, Seibert J, Hakala K, Bock A, Wagener T, *et al.* 2015. Accelerating advances in continental domain hydrologic modeling. *Water Resources Research* **51** (12): 10078–10091 DOI: 10.1002/2015WR017498

Arnold JG, Srinivasan R, Muttiah RS, Williams JR. 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association* **34** (1): 73–89 DOI: 10.1111/j.1752-1688.1998.tb05961.x

Asadzadeh M, Razavi S, Tolson BA, Fay D. 2014. Pre-emption strategies for efficient multi-

objective optimization: Application to the development of Lake Superior regulation plan. *Environmental Modelling & Software* **54**: 128–141 DOI: 10.1016/J.ENVSOFT.2014.01.005

Bajamgnigni Gbambie AS, Poulin A, Boucher M-A, Arsenault R. 2017. Added value of alternative information in interpolated precipitation datasets for hydrology. *Journal of Hydrometeorology* **18** (1): 247–264 DOI: 10.1175/JHM-D-16-0032.1

Barr AG, van der Kamp G, Black TA, McCaughey JH, Nesic Z. 2012. Energy balance closure at the BERMS flux towers in relation to the water balance of the White Gull Creek watershed 1999–2009. *Agricultural and Forest Meteorology* **153**: 3–13 DOI: 10.1016/j.agrformet.2011.05.017

Beck HE, van Dijk A. I. J. M, Levizzani V, Schellekens J, Miralles DG, Martens B, and de Roo A. 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589–615, https://doi.org/10.5194/hess-21-589-2017.

Bekele EG, Nicklow JW. 2007. Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology* **341** (3–4): 165–176 DOI: 10.1016/j.jhydrol.2007.05.014

Berry P, Yassin F, Belcher K, Lindenschmidt K-E. 2017a. An economic assessment of local farm multi-purpose surface water retention systems in a Canadian Prairie setting. *Applied Water Science* **7** (8): 4461–4478 DOI: 10.1007/s13201-017-0592-7

Berry P, Yassin F, Belcher K, Lindenschmidt K-E. 2017b. An Economic Assessment of Local Farm Multi-Purpose Surface Water Retention Systems under Future Climate Uncertainty. *Sustainability* **9** (3): 456 DOI: 10.3390/su9030456

Berry P, Yassin F, Grosshans R, Lindenschmidt K-E. 2017c. Surface water retention systems for cattail production as a biofuel. *Journal of Environmental Management* **203**: 500–509 DOI: 10.1016/j.jenvman.2017.08.019

Best MJ, Pryor M, Clark DB, Rooney GG, Essery R. LH, Ménard CB, Edwards JM, Hendry MA, Porson A, Gedney N, *et al.* 2011. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development* **4** (3): 677–699 DOI: 10.5194/gmd-4-677-2011

Beven K, Binley A. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* **6** (3): 279–298 DOI: 10.1002/hyp.3360060305

Beven K, Cloke H, Pappenberger F, Lamb R, Hunter N. 2015. Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface. *Science China*

Earth Sciences **58** (1): 25–35 DOI: 10.1007/s11430-014-5003-4

Beven K. 1990. A Discussion of Distributed Hydrological Modelling.Springer, Dordrecht; 255–278. DOI: 10.1007/978-94-009-0257-2_13

Beven K. 2012. *Rainfall-runoff modelling : the primer*. Wiley-Blackwell. Available at: https://www.wiley.com/en-us/Rainfall+Runoff+Modelling%3A+The+Primer%2C+2nd+Edition-p-9780470714591

Beven KJ, Kirkby MJ, Schofield N, Tagg AF. 1984. Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments. *Journal of Hydrology* **69** (1–4): 119–143 DOI: 10.1016/0022-1694(84)90159-8

Biancamaria S, Lettenmaier DP, Pavelsky TM. 2016. The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics* **37** (2): 307–337 DOI: 10.1007/s10712-015-9346-y

Biemans H, Haddeland I, Kabat P, Ludwig F, Hutjes RWA, Heinke J, von Bloh W, Gerten D. 2011. Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research* **47** (3): W03509 DOI: 10.1029/2009WR008929

Bierkens MFP, Bell VA, Burek P, Chaney N, Condon LE, David CH, de Roo A, Döll P, Drost N, Famiglietti JS, *et al.* 2015. Hyper-resolution global hydrological modelling: What is next?: 'Everywhere and locally relevant' M. F. P. Bierkens *et al.* Invited Commentary. *Hydrological Processes* **29** (2): 310–320 DOI: 10.1002/hyp.10391

Bondeau A, Smith PC, Zaehle S, Schaphoff S, Lucht W, Cramer W, Gerten D, Lotze-CAMPEN H, MüllER C, Reichstein M, *et al.* 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology* **13** (3): 679–706 DOI: 10.1111/j.1365-2486.2006.01305.x

Burek, P., van der Knijff, J., de Roo A. 2013. LISFLOOD DistributedWater Balance and Flood Simulation Model e Revised User Manual 2013,. JRC Technical Reports. Joint Research Centre of the European Commission. Luxembourg.

Busker T, de Roo A, Gelati E, Schwatke C, Adamovic M, Bisselink B, Pekel J-F, Cottam A. 2019. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences* **23** (2): 669–690 DOI: 10.5194/hess-23-669-2019

Canadian Hydraulics Centre. 2010. Green Kenue Reference Manual. National Research Council, Ottawa, Ont.

Castelletti A, Galelli S, Restelli M, Soncini-Sessa R. 2010. Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research* **46** (9) L04405 DOI: 10.1029/2009WR008898

Chang L-C, Chang F-J, Wang K-W, Dai S-Y. 2010. Constrained genetic algorithms for optimizing multi-use reservoir operation. *Journal of Hydrology* **390** (1–2): 66–74 DOI: 10.1016/J.JHYDROL.2010.06.031

Christensen NS, Wood AW, Voisin N, Lettenmaier DP, Palmer RN. 2004. The effects of climate change on the hydrology and water resources of the colorado river basin. *Climatic Change* **62** (1–3): 337–363 DOI: 10.1023/B:CLIM.0000013684.13621.1f

Clark MP, Fan Y, Lawrence DM, Adam JC, Bolster D, Gochis DJ, Hooper RP, Kumar M, Leung LR, Mackay DS, *et al.* 2015. Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research* **51** (8): 5929–5956 DOI: 10.1002/2015WR017096

Clark MP, Schaefli B, Schymanski SJ, Samaniego L, Luce CH, Jackson BM, Freer JE, Arnold JR, Moore RD, Istanbulluoglu E, *et al.* 2016. Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research* **52** (3): 2350–2365 DOI: 10.1002/2015WR017910

Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV., Wagener T, Hay LE. 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research* **44** (12): W00B02 DOI: 10.1029/2007WR006735

Clark MP, Slater AG. 2006. Probabilistic Quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology* **7** (1): 3–22 DOI: 10.1175/JHM474.1

Coerver HM, Rutten MM, van de Giesen NC. 2018. Deduction of reservoir operating rules for application in global hydrological models. *Hydrology and Earth System Sciences* **22** (1): 831–851 DOI: 10.5194/hess-22-831-2018

Cosby BJ, Hornberger GM, Clapp RB, Ginn TR. 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research* **20** (6): 682–690 DOI: 10.1029/WR020i006p00682

Côté J, Desmarais J-G, Gravel S, Méthot A, Patoine A, Roch M, Staniforth A. 1998. The Operational CMC–MRB Global Environmental Multiscale (GEM) Model. Part II: Results. *Monthly Weather Review* **126** (6): 1397–1418 DOI: 10.1175/1520-

0493(1998)126<1397:TOCMGE>2.0.CO;2

Crow WT, Wood EF, Pan M. 2003. Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals. *Journal of Geophysical Research* **108** (D23): 4725 DOI: 10.1029/2002JD003292

d'Orgeval T, Polcher J, de Rosnay P. 2008. Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes. *Hydrology and Earth System Sciences* **12** (6): 1387–1401 DOI: 10.5194/hess-12-1387-2008

Dane JH, Vrugt JA, Unsal E. 2011. Soil hydraulic functions determined from measurements of air permeability, capillary modeling, and high-dimensional parameter estimation. *Vadose Zone Journal* **10** (1): 459 DOI: 10.2136/vzj2010.0053

Davison B, Pietroniro A, Fortin V, Leconte R, Mamo M, Yau MK, Davison B, Pietroniro A, Fortin V, Leconte R, *et al.* 2016. What is missing from the prescription of hydrology for land surface schemes? *Journal of Hydrometeorology* **17** (7): 2013–2039 DOI: 10.1175/JHM-D-15-0172.1

Davison B, Pohl S, Domes P, Marsh P, Pietroniro A., MacKay M. 2006. Characterizing snowmelt variability in a land-surface-hydrologic model. *Atmosphere-Ocean* **44** (3): 271–287 DOI: 10.3137/ao.440305

de Rosnay P. 2003. Integrated parameterization of irrigation in the land surface model ORCHIDEE. Validation over Indian Peninsula. *Geophysical Research Letters* **30** (19): 1986 DOI: 10.1029/2003GL018024

Degu AM, Hossain F, Niyogi D, Pielke R, Shepherd JM, Voisin N, Chronis T. 2011. The influence of large dams on surrounding climate and precipitation patterns. *Geophysical Research Letters* **38** (4): n/a-n/a DOI: 10.1029/2010GL046482

Dibike YB, Coulibaly P. 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of Hydrology* **307** (1–4): 145–163 DOI: 10.1016/J.JHYDROL.2004.10.012

Döll P, Douville H, Güntner A, Müller Schmied H, Wada Y. 2016. Modelling Freshwater Resources at the Global Scale: Challenges and Prospects. *Surveys in Geophysics* **37** (2): 195–221 DOI: 10.1007/s10712-015-9343-1

Döll P, Fiedler K, Zhang J. 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences* **13** (12): 2413–2432 DOI: 10.5194/hess-13-2413-2009

Döll P, Kaspar F, Lehner B. 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology* **270** (1–2): 105–134 DOI: 10.1016/S0022-1694(02)00283-4

Döll P, Siebert S. 2002. Global modeling of irrigation water requirements. *Water Resources Research* **38** (4): 8-1-8–10 DOI: 10.1029/2001WR000355

Duan Q, Andréassian V, Franks S, Goteti G, Gupta HV, Gusev YM, Habets F, Hall A, Hay L, Hogue T, *et al.* 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology* **320** (1–2): 3–17 DOI: 10.1016/J.JHYDROL.2005.07.031

Duan Q, Sorooshian S, Gupta V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28** (4): 1015–1031 DOI: 10.1029/91WR02985

Eagleson PS. 1986. The emergence of global-scale hydrology. *Water Resources Research* **22** (9S): 6S-14S DOI: 10.1029/WR022i09Sp0006S

Ebert EE, Janowiak JE, Kidd C. 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society* **88** (1): 47–64 DOI: 10.1175/BAMS-88-1-47

Ehsani N, Fekete BM, Vörösmarty CJ, Tessler ZD. 2016. A neural network based general reservoir operation scheme. *Stochastic Environmental Research and Risk Assessment* **30** (4): 1151–1166 DOI: 10.1007/s00477-015-1147-9

Ehsani N, Vörösmarty CJ, Fekete BM, Stakhiv EZ. 2017. Reservoir operations under climate change: Storage capacity options to mitigate risk. *Journal of Hydrology* **555**: 435–446 DOI: 10.1016/J.JHYDROL.2017.09.008

Eum H-I, Dibike Y, Prowse T, Bonsal B. 2014. Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada. *Hydrological Processes* **28** (14): 4250–4271 DOI: 10.1002/hyp.10236

Fang X, Pomeroy JW. 2007. Snowmelt runoff sensitivity analysis to drought on the Canadian prairies. *Hydrological Processes* **21**: 2594–2609 DOI: 10.1002/hyp.6796

Faramarzi M, Abbaspour KC, Adamowicz WL, Lu W, Fennell J, Zehnder AJB, Goss GG. 2016. Uncertainty based assessment of dynamic freshwater scarcity in semi-arid watersheds of Alberta, Canada. *Journal of Hydrology: Regional Studies* **9**: 48–68 DOI: 10.1016/j.ejrh.2016.11.003

Fortin V, Roy G, Donaldson N, Mahidjiba A. 2015. Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *Journal of Hydrology* **531**: 296–307 DOI: 10.1016/J.JHYDROL.2015.08.003

Fortin V, Roy G, Stadnyk T, Koenig K, Gasset N, Mahidjiba A. 2018. Ten years of science based on the Canadian Precipitation Analysis: A CaPA system overview and literature review. *Atmosphere-Ocean* **56** (3): 178–196 DOI: 10.1080/07055900.2018.1474728

Fraternali P, Castelletti A, Soncini-Sessa R, Vaca Ruiz C, Rizzoli AE. 2012. Putting humans in the loop: Social computing for water resources management. *Environmental Modelling & Software* **37**: 68–77 DOI: 10.1016/J.ENVSOFT.2012.03.002

Freeze RA, Harlan RL. 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology* **9** (3): 237–258 DOI: 10.1016/0022-1694(69)90020-1

Gao H, Birkett C, Lettenmaier DP. 2012. Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research* **48** (9): W09504 DOI: 10.1029/2012WR012063

Garambois P-A, Monnier J. 2015. Inference of effective river properties from remotely sensed observations of water surface. *Advances in Water Resources* **79**: 103–120 DOI: 10.1016/J.ADVWATRES.2015.02.007

Gosling SN, Arnell NW. 2011. Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. *Hydrological Processes* **25** (7): 1129–1145 DOI: 10.1002/hyp.7727

Gudmundsson L, Tallaksen LM, Stahl K, Clark DB, Dumont E, Hagemann S, Bertrand N, Gerten D, Heinke J, Hanasaki N, *et al.* 2012. Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe. *Journal of Hydrometeorology* **13** (2): 604–620 DOI: 10.1175/JHM-D-11-083.1

Guillaume JHA, Jakeman JD, Marsili-Libelli S, Asher M, Brunner P, Croke B, Hill MC, Jakeman AJ, Keesman KJ, Razavi S, *et al.* 2019. Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software* DOI: 10.1016/J.ENVSOFT.2019.07.007

Guo X, Hu T, Zeng X, Li X. 2013. Extension of parametric rule with the hedging rule for managing multireservoir system during droughts. *Journal of Water Resources Planning and Management* **139** (2): 139–148 DOI: 10.1061/(ASCE)WR.1943-5452.0000241

Gupta H V., Kling H, Yilmaz KK, Martinez GF. 2009. Decomposition of the mean squared error

and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2): 80–91 DOI: 10.1016/J.JHYDROL.2009.08.003

Gupta HV, Sorooshian S, Yapo PO. 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research* **34** (4): 751–763 DOI: 10.1029/97WR03495

Gupta HV., Bastidas LA, Sorooshian S, Shuttleworth WJ, Yang ZL. 1999. Parameter estimation of a land surface scheme using multicriteria methods. *Journal of Geophysical Research: Atmospheres* **104** (D16): 19491–19503 DOI: 10.1029/1999JD900154

Gupta HV., Razavi S. 2018. Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research* **54** (11): 8692–8717 DOI: 10.1029/2018WR022668

Gupta HV., Wagener T, Liu Y. 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes* **22** (18): 3802–3813 DOI: 10.1002/hyp.6989

Haddeland I, Heinke J, Biemans H, Eisner S, Flörke M, Hanasaki N, Konzmann M, Ludwig F, Masaki Y, Schewe J, *et al.* 2014. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences of the United States of America* **111** (9): 3251–6 DOI: 10.1073/pnas.1222475110

Haddeland I, Skaugen T, Lettenmaier DP. 2006. Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters* **33** (8): L08406 DOI: 10.1029/2006GL026047

Hadka D, Reed P. 2013. BORG: an auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation* **21** (2): 231–259 DOI: 10.1162/EVCO_a_00075

Hadka D, Reed P. 2015. Large-scale parallelization of the BORG multiobjective evolutionary algorithm to enhance the management of complex environmental systems. *Environmental Modelling & Software* **69**: 353-369 DOI: 10.1016/j.envsoft.2014.10.014

Haghnegahdar A, Razavi S, Yassin F, Wheater H. 2017. Multi-criteria sensitivity analysis as a diagnostic tool for understanding model behavior and characterizing model uncertainty. *Hydrological Processes* **31** (25): 4462–4476 DOI: 10.1002/hyp.11358

Haghnegahdar A, Tolson BA, Davison B, Seglenieks FR, Klyszejko E, Soulis ED, Fortin V, Matott LS. 2014. Calibrating Environment Canada's MESH modelling system over the Great Lakes Basin. *Atmosphere-Ocean* **52** (4): 281–293 DOI: 10.1080/07055900.2014.939131

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y, Tanaka K. 2008a. An

integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences* **12** (4): 1007–1025 DOI: 10.5194/hess-12-1007-2008

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y, Tanaka K. 2008b. An integrated model for the assessment of global water resources – Part 2: Applications and assessments. *Hydrology and Earth System Sciences* **12** (4): 1027–1037 DOI: 10.5194/hess-12-1027-2008

Hanasaki N, Kanae S, Oki T. 2006. A reservoir operation scheme for global river routing models. *Journal of Hydrology* **327** (1–2): 22–41 DOI: 10.1016/j.jhydrol.2005.11.011

Hossain F, Degu AM, Yigzaw W, Burian S, Niyogi D, Shepherd JM, Pielke R. 2012. Climate feedback–based provisions for dam design, operations, and water management in the 21st century. *Journal of Hydrologic Engineering* **17** (8): 837–850 DOI: 10.1061/(ASCE)HE.1943-5584.0000541

Houser PR, Gupta HV., Shuttleworth WJ, Famiglietti JS. 2001. Multiobjective calibration and sensitivity of a distributed land surface water and energy balance model. *Journal of Geophysical Research: Atmospheres* **106** (D24): 33421–33433 DOI: 10.1029/2000JD900803

Hsu KL, Behrangi A, Imam B, Sorooshian S. 2010. Extreme Precipitation Estimation Using Satellite-Based PERSIANN-CCS Algorithm. *Satellite Rainfall Applications for Surface Hydrology*: 49-67. DOI:10.1007/978-90-481-2915-7_4

Huang J, Halpenny J, van der Wal W, Klatt C, James TS, Rivera A. 2012. Detectability of groundwater storage change within the Great Lakes Water Basin using GRACE. *Journal of Geophysical Research: Solid Earth* **117** (B8): B08401 DOI: 10.1029/2011JB008876

Huffman GJ, Adler RF, Stocker E, Bolvin DT, Nelkin EJ. 2002. Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time

Huffman, GJ. et al., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1): 38-55

Islam Z, Gan TY. 2014. Effects of climate change on the surface-water management of the south saskatchewan river basin. *Journal of Water Resources Planning and Management* **140** (3): 332–342 DOI: 10.1061/(ASCE)WR.1943-5452.0000326

Kasprzyk JR, Nataraj S, Reed PM, Lempert RJ. 2013. Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*

42: 55–71 DOI: 10.1016/J.ENVSOFT.2012.12.007

Kirchherr J, Pohlner H, Charles KJ. 2016. Cleaning up the big muddy: A meta-synthesis of the research on the social impact of dams. *Environmental Impact Assessment Review* **60**: 115–125 DOI: 10.1016/J.EIAR.2016.02.007

Kobayashi, S. et al., 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. J Meteorol Soc Jpn, 93(1): 5-48. DOI:10.2151/jmsj.2015-001

Kouwen N, Soulis ED, Pietroniro A, Donald J, Harrington RA. 1993. Grouped Response Units for distributed hydrologic modeling. *Journal of Water Resources Planning and Management* **119** (3): 289–305 DOI: 10.1061/(ASCE)0733-9496(1993)119:3(289)

Kuczera G, Mroczkowski M. 1998. Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resources Research* **34** (6): 1481–1489 DOI: 10.1029/98WR00496

Labadie JW. 2004. Optimal operation of multireservoir systems: state-of-the-art review. *Journal of Water Resources Planning and Management* **130** (2): 93–111 DOI: 10.1061/(ASCE)0733-9496(2004)130:2(93)

Lambert A, Huang J, van der Kamp G, Henton J, Mazzotti S, James TS, Courtier N, Barr AG. 2013. Measuring water accumulation rates using GRACE data in areas experiencing glacial isostatic adjustment: The Nelson River basin. *Geophysical Research Letters* **40** (23): 6118–6122 DOI: 10.1002/2013GL057973

Landerer FW, Swenson SC. 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research* **48** (4): W04531 DOI: 10.1029/2011WR011453

Layek G. 2015. *Introduction to Dynamical Systems and Chaos*. Springer India: West Bengal, India. Available at: http://link.springer.com/content/pdf/10.1007/978-81-322-2556-0.pdf [Accessed 11 February 2016]

Lehner B, Döll P. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* **296** (1): 1–22 DOI: 10.1016/j.jhydrol.2004.03.028

Lehner B, Liermann CR, Revenga C, Vörösmarty C, Fekete B, Crouzet P, Döll P, Endejan M, Frenken K, Magome J, *et al.* 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment* **9** (9): 494–502 DOI: 10.1890/100125

Liang X, Lettenmaier DP, Wood EF, Burges SJ. 1994. A simple hydrologically based model of

land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research* **99** (D7): 14415-14428 DOI: 10.1029/94JD00483

Liebe J, van de Giesen N, Andreini M. 2005. Estimation of small reservoir storage capacities in a semi-arid environment: A case study in the Upper East Region of Ghana. *Physics and Chemistry of the Earth, Parts A/B/C* **30** (6–7): 448–454 DOI: 10.1016/J.PCE.2005.06.011

Liu Y, Gupta HV. 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research* **43** (7): W07401 DOI: 10.1029/2006WR005756

Livneh B, Lettenmaier DP. 2012. Multi-criteria parameter estimation for the Unified Land Model. *Hydrology and Earth System Sciences* **16** (8): 3029–3048 DOI: 10.5194/hess-16-3029-2012

Lo M-H, Famiglietti JS, Yeh PJ-F, Syed TH. 2010. Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. *Water Resources Research* **46** (5): W05517 DOI: 10.1029/2009WR007855

Long D, Longuevergne L, Scanlon BR. 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research* **50** (2): 1131–1151 DOI: 10.1002/2013WR014581

Longuevergne L, Wilson CR, Scanlon BR, Crétaux JF. 2013. GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage. *Hydrology and Earth System Sciences* **17** (12): 4817–4830 DOI: 10.5194/hess-17-4817-2013

MacDonald MK, Pomeroy JW, Pietroniro A. 2009. Parameterizing redistribution and sublimation of blowing snow for hydrological models: tests in a mountainous subarctic catchment. *Hydrological Processes* **23** (18): 2570–2583 DOI: 10.1002/hyp.7356

MacKay MD. 2012. A process-oriented small lake scheme for coupled climate modeling applications. *Journal of Hydrometeorology* **13** (6): 1911–1924 DOI: 10.1175/JHM-D-11-0116.1

Madsen H. 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources* **26** (2): 205–216 DOI: 10.1016/S0309-1708(02)00092-1

Maggioni V, Sapiano M.R.P, Adler R.F, Tian Y.D, Huffman G.J. 2014. An Error Model for Uncertainty Quantification in High-Time-Resolution Precipitation Products. J Hydrometeorol, 15(3): 1274-1292. DOI:10.1175/Jhm-D-13-0112.1

Mahfouf J-F, Brasnett B, Gagnon S. 2007. A Canadian Precipitation Analysis (CaPA) project: description and preliminary results. *Atmosphere-Ocean* **45** (1): 1–17 DOI: 10.3137/ao.v450101

Mahfouf J-F, Brasnett B, Gagnon S. 2007. A Canadian Precipitation Analysis (CaPA) Project: description and preliminary results. *Atmosphere-Ocean* **45** (1): 1–17 DOI: 10.3137/ao.v450101

Maier HR, Dandy GC. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software* **15** (1): 101–124 DOI: 10.1016/S1364-8152(99)00007-9

Maier HR, Razavi S, Kapelan Z, Matott LS, Kasprzyk J, Tolson BA. 2019. Introductory overview: Optimization using evolutionary algorithms and other metaheuristics. *Environmental Modelling & Software* **114**: 195–213 DOI: 10.1016/J.ENVSOFT.2018.11.018

Manabe S. 1969. Climate and the ocean circulation. *Monthly Weather Review* **97** (11): 739–774 DOI: 10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2

Meigh JR, McKenzie AA, Sene KJ. 1999. A grid-based approach to water scarcity estimates for eastern and southern africa. *Water Resources Management* **13** (2): 85–115 DOI: 10.1023/A:1008025703712

Mekis E, Hogg WD. 1999. Rehabilitation and analysis of Canadian daily precipitation time series. *Atmosphere-Ocean* **37** (1): 53–85 DOI: 10.1080/07055900.1999.9649621

Mekis É, Vincent LA. 2011. An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean* **49** (2): 163–177 DOI: 10.1080/07055900.2011.583910

Mekonnen MA., Wheater HS, Ireson AM, Spence C, Davison B, Pietroniro A. 2014. Towards an improved land surface scheme for prairie landscapes. *Journal of Hydrology* **511**: 105–116 DOI: 10.1016/j.jhydrol.2014.01.020

Melsen LA, Teuling JA, Torfs JJFP, Uijlenhoet R, Mizukami N, Clark MP. 2016. HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models. *Hydrology and Earth System Sciences* **20** (3): 1069–1079 DOI: 10.5194/hess-20-1069-2016

Mendoza PA, Clark MP, Barlage M, Rajagopalan B, Samaniego L, Abramowitz G, Gupta H. 2015. Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research* **51** (1): 716–728 DOI: 10.1002/2014WR015820

Mengistu SG, Ali MA, Yassin F. 2016. Assessment of the Sensitivity of Streamflow Simulations to Changes in Patch Resolution Using GIS Based Hydro-Ecologic Model. *Open Journal of Modern Hydrology* **06** (02): 66–78 DOI: 10.4236/ojmh.2016.62007

Mengistu SG, Spence C. 2016. Testing the ability of a semidistributed hydrological model to simulate contributing area. *Water Resources Research* **52** (6): 4399–4415 DOI: 10.1002/2016WR018760

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, *et al.* 2006. North American Regional Reanalysis. *Bulletin of the American Meteorological Society* **87** (3): 343–360 DOI: 10.1175/BAMS-87-3-343

Middelkoop H, Daamen K, Gellens D, Grabs W, Kwadijk JCJ, Lang H, Parmet BWAH, Schädler B, Schulla J, Wilke K. 2001. Impact of climate change on hydrological regimes and water resources management in the Rhine Basin. *Climatic Change* **49** (1/2): 105–128 DOI: 10.1023/A:1010784727448

Mizukami N, P. Clark M. 2014. Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. *Journal of Hydrometeorology* **15** (1): 474–488 DOI: 10.1175/JHM-D-13-036.1

Moore RJ. 2007. The PDM rainfall-runoff model. *Hydrology and Earth System Sciences* **11** (1): 483–499 DOI: 10.5194/hess-11-483-2007

Morris MD. 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* **33** (2): 161–174 DOI: 10.1080/00401706.1991.10484804

Mu Q, Heinsch FA, Zhao M, Running SW. 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment* **111** (4): 519–536 DOI: 10.1016/J.RSE.2007.04.015

Müller Schmied H, Eisner S, Franz D, Wattenbach M, Portmann FT, Flörke M, Döll P. 2014. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences* **18** (9): 3511–3538 DOI: 10.5194/hess-18-3511-2014

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* **10** (3): 282–290 DOI: 10.1016/0022-1694(70)90255-6

Nasonova ON, Gusev YM, Kovalev YE. 2009. Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: a case study using MOPEX

materials. *Journal of Hydrometeorology* **10**: 1128–1150 DOI: 10.2307/24912043

Nazemi A, Wheater HS. 2015a. On inclusion of water resource management in Earth system models; Part 1: Problem definition and representation of water demand. *Hydrology and Earth System Sciences* **19** (1): 33–61 DOI: 10.5194/hess-19-33-2015

Nazemi A, Wheater HS. 2015b. On inclusion of water resource management in Earth system models; Part 2: Representation of water supply and allocation and opportunities for improved modeling. *Hydrology and Earth System Sciences* **19** (1): 63–90 DOI: 10.5194/hess-19-63-2015

Neitsch SLL, Arnold JGG, Kiniry JRR, Srinivasan R, Williams JRR. 2011. Soil and Water Assessment Tool Input Output File Documentation Version 2009, Texas Water resources institute technical report, Texas.

Nilsson C, Reidy CA, Dynesius M, Revenga C. 2005. Fragmentation and flow regulation of the world's large river systems. *Science* **308** (5720): 405-408 DOI: 10.1126/science.1107887

Oki T, Kanae S. 2006. Global hydrological cycles and world water resources. *Science (New York, N.Y.)* **313** (5790): 1068–1072 DOI: 10.1126/science.1128845

Oleson KW, Lawrence DM, Bonan G, Flanner MG, Kluzek E, Lawrence PJ, Levis S, Swenson SC, Thornton E, *et al.* 2010. Technical description of version 5.0 of the Community Land Model (CLM) available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.172.7769

Onogi, K. et al., 2007. The JRA-25 reanalysis. J Meteorol Soc Jpn, 85(3): 369-432. DOI:DOI 10.2151/jmsj.85.369

Overgaard J, Rosbjerg D, Butts MB. 2006. Land-surface modelling in hydrological perspective – a review. *Biogeosciences* **3** (2): 229–241 DOI: 10.5194/bg-3-229-2006

Ozdogan M, Rodell M, Beaudoing HK, Toll DL, Ozdogan M, Rodell M, Beaudoing HK, Toll DL. 2010. Simulating the effects of irrigation over the united states in a land surface model based on satellite-derived agricultural data. *Journal of Hydrometeorology* **11** (1): 171–184 DOI: 10.1175/2009JHM1116.1

Pechlivanidis IG, Jackson B, McMillan H, Gupta H. 2014. Use of an entropy-based metric in multiobjective calibration to improve model performance. *Water Resources Research* **50** (10): 8066–8083 DOI: 10.1002/2013WR014537

Pietroniro A, Fortin V, Kouwen N, Neal C, Turcotte R, Davison B, Verseghy D, Soulis ED,

Caldwell R, Evora N, *et al.* 2007. Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences* **11** (4): 1279–1294 DOI: 10.5194/hess-11-1279-2007

Pietroniro A, Soulis ED. 2003. A hydrology modelling framework for the Mackenzie GEWEX programme. *Hydrological Processes* **17** (3): 673–676 DOI: 10.1002/hyp.5104

Pitman AJ. 2003. The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology* **23** (5): 479–510 DOI: 10.1002/joc.893

Pohl S, Marsh P. 2006. Modelling the spatial–temporal variability of spring snowmelt in an arctic catchment. *Hydrological Processes* **20** (8): 1773–1792 DOI: 10.1002/hyp.5955

Pokhrel Y, Hanasaki N, Koirala S, Cho J, Yeh PJ-F, Kim H, Kanae S, Oki T, Pokhrel Y, Hanasaki N, *et al.* 2012. Incorporating anthropogenic water regulation modules into a land surface model. Journal of Hydrometerology **13** (1): 255-269 DOI: 10.1175/JHM-D-11-013.1

Pokhrel YN, Felfelani F, Shin S, Yamada TJ, Satoh Y. 2017. Modeling large-scale human alteration of land surface hydrology and climate. *Geoscience Letters* **4** (1): 10 DOI: 10.1186/s40562-017-0076-5

Pokhrel YN, Hanasaki N, Wada Y, Kim H. 2016. Recent progresses in incorporating human land-water management into global land surface models toward their integration into Earth system models. *Wiley Interdisciplinary Reviews: Water* **3** (4): 548–574 DOI: 10.1002/wat2.1150

Pomeroy J, Fang X, Westbrook C, Minke A, Guo X, Brown T. 2010. Prairie Hydrological Model Study Final Report. Saskatoon. Available at: http://www.usask.ca/hydrology/reports/CHRpt07_PHMS-Final-Report_Jan10.pdf

Qiao L, Herrmann RB, Pan Z. 2013. Parameter uncertainty reduction for SWAT Using GRACE, streamflow, and groundwater table data for Lower Missouri River Basin. *JAWRA Journal of the American Water Resources Association* **49** (2): 343–358 DOI: 10.1111/jawr.12021

Raad D, Sinske A, van Vuuren J. 2009. Robust multi-objective optimization for water distribution system design using a meta-metaheuristic. *International Transactions in Operational Research* **16** (5): 595–626 DOI: 10.1111/j.1475-3995.2009.00705.x

Rakovec O, Kumar R, Mai J, Cuntz M, Thober S, Zink M, Attinger S, Schäfer D, Schrön M, Samaniego L, *et al.* 2015. Multiscale and multivariate evaluation of water fluxes and states over European River Basins. *Journal of Hydrometeorology* **17** (1): 287–307 DOI: 10.1175/JHM-D-15-0054.1

Razavi S, Asadzadeh M, Tolson B, Fay D, Moin S, Bruxer J, Fan Y. 2014. Evaluation of new control structures for regulating the great lakes system: multiscenario, multireservoir optimization approach. *Journal of Water Resources Planing and Managmenet* **140** (8): 04014018 DOI: 10.1061/(ASCE)WR.1943-5452.0000375

Razavi S, Asce AM, Asadzadeh M, Asce AM, Tolson B, Fay D, Moin S, Bruxer J, Fan Y. 2014. Evaluation of new control structures for regulating the great lakes system: multiscenario, multireservoir optimization approach DOI: 10.1061/(ASCE)WR.1943-5452.0000375

Razavi S, Elshorbagy A, Wheater H, Sauchyn D. 2015. Toward understanding nonstationarity in climate and hydrology through tree ring proxy records. *Water Resources Research* **51** (3): 1813–1830 DOI: 10.1002/2014WR015696

Razavi S, Gupta H V. 2019. A multi-method Generalized Global Sensitivity Matrix approach to accounting for the dynamical nature of earth and environmental systems models. *Environmental Modelling & Software* **114**: 1–11 DOI: 10.1016/J.ENVSOFT.2018.12.002

Razavi S, Gupta HV. 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of 'global' sensitivity in Earth and Environmental systems models. *Water Resources Research* **51** (5): 3070–3092 DOI: 10.1002/2014WR016527

Razavi S, Gupta HV. 2016a. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research* **52** (1): 423–439 DOI: 10.1002/2015WR017558

Razavi S, Gupta HV. 2016b. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research* **52** (1): 440–455 DOI: 10.1002/2015WR017559

Razavi S, Karamouz M. 2007. Adaptive neural networks for flood routing in river systems. *Water International* **32** (3): 360–375 DOI: 10.1080/02508060708692216

Razavi S, Tolson BA, Burn DH. 2012. Review of surrogate modeling in water resources. *Water Resources Research* **48** (7): W07401 DOI: 10.1029/2011WR011527

Razavi S, Tolson BA, Matott LS, Thomson NR, MacLean A, Seglenieks FR. 2010. Reducing the computational cost of automatic calibration through model preemption. *Water Resources Research* **46** (11) DOI: 10.1029/2009WR008957

Razavi S, Tolson BA. 2011. A New Formulation for Feedforward Neural Networks. *IEEE Transactions on Neural Networks* **22** (10): 1588–1598 DOI: 10.1109/TNN.2011.2163169

Razavi S, Tolson BA. 2013. An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research* **49** (12): 8418–8431 DOI: 10.1002/2012WR013442

Ricard S, Bourdillon R, Roussel D, Turcotte R. 2013. Global calibration of distributed hydrological models for large-scale applications. *Journal of Hydrologic Engineering* **18** (6): 719–721 DOI: 10.1061/(ASCE)HE.1943-5584.0000665

Rienecker M.M. et al., 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. J Climate, 24(14): 3624-3648. DOI:10.1175/Jcli-D-11-00015.1

Rost S, Gerten D, Bondeau A, Lucht W, Rohwer J, Schaphoff S. 2008. Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research* **44** (9): W09405 DOI: 10.1029/2007WR006331

Sapriza-Azuri G, Gamazo P, Razavi S, Wheater HS. 2018. On the appropriate definition of soil profile configuration and initial conditions for land surface–hydrology models in cold regions. *Hydrology and Earth System Sciences* **22** (6): 3295–3309 DOI: 10.5194/hess-22-3295-2018

Savtchenko A, Ouzounov D, Ahmad S, Acker J, Leptoukh G, Koziana J, Nickless D. 2004. Terra and Aqua MODIS products available from NASA GES DAAC. *Advances in Space Research* **34** (4): 710–714 DOI: 10.1016/j.asr.2004.03.012

Sawicz K, Wagener T, Sivapalan M, Troch PA, Carrillo G. 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences* **15** (9): 2895–2911 DOI: 10.5194/hess-15-2895-2011

Schirmer M, Jamieson B. 2015. Verification of analysed and forecasted winter precipitation in complex terrain. *The Cryosphere* **9** (2): 587–601 DOI: 10.5194/tc-9-587-2015

Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research* **38** (11): 23-1-23–14 DOI: 10.1029/2001WR000978

Seo K-W, Wilson CR, Famiglietti JS, Chen JL, Rodell M. 2006. Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research* **42** (5): W05417 DOI: 10.1029/2005WR004255

Sheffield J, Goteti G, Wood EF, Sheffield J, Goteti G, Wood EF. 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate* **19** (13): 3088–3111 DOI: 10.1175/JCLI3790.1

Sheikholeslami R, Yassin F, Lindenschmidt K-E, Razavi S. 2017. Improved Understanding of River Ice Processes Using Global Sensitivity Analysis Approaches. *Journal of Hydrologic Engineering* **22** (11): 04017048 DOI: 10.1061/(ASCE)HE.1943-5584.0001574

Sheikholeslami R, Razavi S, Gupta HV., Becker W, Haghnegahdar A. 2019. Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. *Environmental Modelling & Software* **111**: 282–299 DOI: 10.1016/J.ENVSOFT.2018.09.002

Shen S.S.P, Dzikowski P, Li G.L, Griffith D. 2001. Interpolation of 1961-97 daily temperature and precipitation data onto Alberta polygons of ecodistrict and soil landscapes of Canada. J Appl Meteorol, 40(12): 2162-2177. DOI:Doi 10.1175/1520-0450(2001)040<2162:Iodtap>2.0.Co;2

Shook K, Pomeroy JW, Spence C, Boychuk L. 2013. Storage dynamics simulations in prairie wetland hydrology models: evaluation and parameterization. *Hydrological Processes* **27** (13): 1875–1889 DOI: 10.1002/hyp.9867

Siebert S, Henrich V, Frenken K, Burke J. 2013. Update of the digital global map of irrigation areas to version 5. Rheinische Friedrich-Wilhelms-Universituy, Bonn, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy: 171 Available at: https://www.lap.uni-bonn.de/research/downloads/gmia/siebert_et_al_2013_gmia5

Silvestro F, Gabellani S, Rudari R, Delogu F, Laiolo P, Boni G. 2015. Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrology and Earth System Sciences* **19** (4): 1727–1751 DOI: 10.5194/hess-19-1727-2015

Sivakumar B. 2012. Socio-hydrology: not a new science, but a recycled and re-worded hydrosociology. *Hydrological Processes* **26** (24): 3788–3790 DOI: 10.1002/hyp.9511

Sivapalan M, Savenije HHG, Blöschl G. 2012. Socio-hydrology: A new science of people and water. *Hydrological Processes* **26** (8): 1270–1276 DOI: 10.1002/hyp.8426

Sobol′ I. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* **55** (1–3): 271–280 DOI: 10.1016/S0378-4754(00)00270-6

Solander KC, Reager JT, Thomas BF, David CH, Famiglietti JS, Solander KC, Reager JT, Thomas BF, David CH, Famiglietti JS. 2016. Simulating human water regulation: the development of an optimal complexity, climate-adaptive reservoir management model for an LSM. *Journal*

of Hydrometeorology **17** (3): 725–744 DOI: 10.1175/JHM-D-15-0056.1

Sood A, Smakhtin V. 2015. Global hydrological models: a review. *Hydrological Sciences Journal* **60** (4): 549–565 DOI: 10.1080/02626667.2014.950580

Soulis ED, Craig JR, Fortin V, Liu G. 2011. A simple expression for the bulk field capacity of a sloping soil horizon. *Hydrological Processes* **25** (1): 112–116 DOI: 10.1002/hyp.7827

Soulis ED, Snelgrove KR, Kouwen N, Seglenieks F, Verseghy DL. 2000. Towards closing the vertical water balance in Canadian atmospheric models: Coupling of the land surface scheme CLASS with the distributed hydrological model watflood. *Atmosphere-Ocean* **38** (1): 251–269 DOI: 10.1080/07055900.2000.9649648

Strobl E, Strobl RO. 2011. The distributional impact of large dams: Evidence from cropland productivity in Africa. *Journal of Development Economics* **96** (2): 432–450 DOI: 10.1016/J.JDEVECO.2010.08.005

Swenson S. 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research* **107** (B9): 2193 DOI: 10.1029/2001JB000576

Takata K, Emori S, Watanabe T. 2003. Development of the minimal advanced treatments of surface interaction and runoff. *Global and Planetary Change* **38** (1–2): 209–222 DOI: 10.1016/S0921-8181(03)00030-4

Tanzeeba S, Gan TY. 2011. Potential impact of climate change on the water availability of South Saskatchewan River Basin. *Climatic Change* **112** (2): 355–386 DOI: 10.1007/s10584-011-0221-7

Tapley BD, Bettadpur S, Ries JC, Thompson PF, Watkins MM. 2004. GRACE measurements of mass variability in the Earth system. *Science (New York, N.Y.)* **305** (5683): 503–505 DOI: 10.1126/science.1099192

Tolson BA, Shoemaker CA. 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* **43** (1): W01413 DOI: 10.1029/2005WR004723

Uppala S.M. et al., 2005. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society, 131(612): 2961-3012.

U.S. Department of Agriulture. 1951. Soil Survey Manual. U.S. Dep. of Agric. Handbk. DOI: 10.1097/00010694-195112000-00022

van Beek LPH, Bierkens MFP. 2009. The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification Available at: https://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf

van Beek LPH, Wada Y, Bierkens MFP. 2011. Global monthly water stress: 1. Water balance and water availability. *Water Resources Research* **47** (7) DOI: 10.1029/2010WR009791

van der Knijff JM, Younis J, De Roo APJ. 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science* **24** (2): 189–212 DOI: 10.1080/13658810802549154

van Dijk AIJM, Renzullo LJ. 2011. Water resource monitoring systems and the role of satellite observations. *Hydrology and Earth System Sciences* **15** (1): 39–55 DOI: 10.5194/hess-15-39-2011

Verseghy D, McFarlane NA, Lazare M. 1993. CLASS—A Canadian land surface scheme for GCMs, II. Vegetation model and coupled runs. *International Journal of Climatology* **13** (4): 347–370 DOI: 10.1002/joc.3370130402

Verseghy D. 1991. Class-A Canadian land surface scheme for GCMS. I. Soil model. *International Journal of Climatology* **11** (2): 111–133 DOI: 10.1002/joc.3370110202

Verseghy D. 2011. The Canadian land surface scheme (version 3.6): Technical documentation (version 1). Environment Canada, 180 pp.

Verseghy DL, Saunders IR, Bowers JD, Huo Z, Bailey WG. 2000. Application of the Canadian land surface scheme (CLASS) to the simulation of energy and water fluxes over alpine tundra. *Atmosphere-Ocean* **38** (1): 37–55 DOI: 10.1080/07055900.2000.9649639

Verseghy, D. L. 2009. CLASS – The Canadian Land Surface Scheme (version 3.4). Environment Canada, Toronto, Ont.

Voisin N, Li H, Ward D, Huang M, Wigmosta M, Leung LR. 2013a. On an improved sub-regional water resources management representation for integration into earth system models. *Hydrology and Earth System Sciences* **17** (9): 3605–3622 DOI: 10.5194/hess-17-3605-2013

Voisin N, Liu L, Hejazi M, Tesfa T, Li H, Huang M, Liu Y, Leung LR. 2013b. One-way coupling of an integrated assessment model and a water resources model: evaluation and implications of future changes over the US Midwest. *Hydrology and Earth System Sciences* **17** (11): 4555–4575 DOI: 10.5194/hess-17-4555-2013

Voisin N, Liu L, Hejazi M, Tesfa T, Li H, Huang M, Liu Y, Leung LR. 2013b. One-way coupling

of an integrated assessment model and a water resources model: evaluation and implications of future changes over the US Midwest. *Hydrology and Earth System Sciences* **17** (11): 4555–4575 DOI: 10.5194/hess-17-4555-2013

Vörösmarty CJ, Federer CA, Schloss AL. 1998. Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. *Journal of Hydrology* **207** (3–4): 147–169 DOI: 10.1016/S0022-1694(98)00109-7

Vörösmarty CJ, Green P, Salisbury J, Lammers RB, Falkenmark M. 2000. Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science* **289** (5477): 284–288 DOI: 10.1126/science.289.5477.284

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich a, Green P, Glidden S, Bunn SE, Sullivan C a, Liermann CR, *et al.* 2010. Global threats to human water security and river biodiversity. *Nature* **467** (7315): 555–561 DOI: 10.1038/nature09440

Vörösmarty CJ, Meybeck M, Fekete B, Sharma K, Green P, Syvitski JP. 2003. Anthropogenic sediment retention: major global impact from registered river impoundments. *Global and Planetary Change* **39** (1–2): 169–190 DOI: 10.1016/S0921-8181(03)00023-7

Vorosmarty CJ, Sharma KP, Fekete BM, Coepland AH, Holden J, Marble J, Lough JA. 1997. The storage and aging of continental runoff in large reservoir systems of the world. *Ambio* **26** (4): 210-219 [online] Available from: http://www.jstor.org/stable/4314590, 1997.

Vrugt JA, Gupta HV., Bastidas LA, Bouten W, Sorooshian S. 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research* **39** (8): 1214 DOI: 10.1029/2002WR001746

Vrugt JA, Robinson BA, Hyman JM. 2009. Self-adaptive multimethod search for global optimization in real-parameter spaces. *IEEE Transactions on Evolutionary Computation* **13** (2): 243–259 DOI: 10.1109/TEVC.2008.924428

Vrugt JA, Robinson BA. 2007. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences of the United States of America* **104** (3): 708–711 DOI: 10.1073/pnas.0610471104

Wada Y, Bierkens MFP, de Roo A, Dirmeyer PA, Famiglietti JS, Hanasaki N, Konar M, Liu J, Müller Schmied H, Oki T, *et al.* 2017. Human–water interface in hydrological modelling: current status and future directions. *Hydrology and Earth System Sciences* **21** (8): 4169–4193 DOI: 10.5194/hess-21-4169-2017

Wagener T, Sivapalan M, Troch P, Woods R. 2007. Catchment classification and hydrologic

similarity. *Geography Compass* 1 (4): 901–931 DOI: 10.1111/j.1749-8198.2007.00039.x

Wagner W, Blöschl G, Pampaloni P, Calvet J-C, Bizzarri B, Wigneron J-P, Kerr Y. 2007. Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Hydrology Research* **38** (1): 1–20 DOI: 10.2166/nh.2007.029

Wallner M, Haberlandt U, Dietrich J. 2012. Evaluation of different calibration strategies for large scale continuous hydrological modelling. *Advances in Geosciences* **31**: 67–74 DOI: 10.5194/adgeo-31-67-2012

Wang S, Yang Y, Luo Y, Rivera A. 2013. Spatial and seasonal variations in evapotranspiration over Canada's landmass. *Hydrology and Earth System Sciences* **17** (9): 3561–3575 DOI: 10.5194/hess-17-3561-2013

Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P. 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*: 7505–7514 DOI: 10.1002/2014WR015638

Wen L, Lin CA, Wu Z, Lu G, Pomeroy J, Zhu Y. 2011. Reconstructing sixty year (1950-2009) daily soil moisture over the Canadian Prairies using the Variable Infiltration Capacity model. *Canadian Water Resources Journal* **36** (1): 83–102 DOI: 10.4296/cwrj3601083

Werth S, Güntner A, Petrovic S, Schmidt R. 2009. Integration of GRACE mass variations into a global hydrological model. *Earth and Planetary Science Letters* **277** (1–2): 166–173 DOI: 10.1016/j.epsl.2008.10.021

Werth S, Güntner A. 2010. Calibration analysis for water storage variability of the global hydrological model WGHM. *Hydrology and Earth System Sciences* **14** (1): 59–78 DOI: 10.5194/hess-14-59-2010

Wheater H, Gober P. 2013. Water security in the Canadian Prairies: science and management challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371** (2002): 20120409–20120409 DOI: 10.1098/rsta.2012.0409

Wheater H. 2013. Saskatchewan River Basin Regional Hydroclimate Project - *Global Energy and Water Exchange (GEWEX) Newsletter* Available at: http://www.gewex.org/gewex-content/files_mf/1432213474May2013.pdf

Wheater HS, Gober P. 2015. Water security and the science agenda. *Water Resources Research* **51** (7): 5406–5424 DOI: 10.1002/2015WR016892

Wisser D, Fekete BM, Vörösmarty CJ, Schumann AH. 2010. Reconstructing 20th century global

hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H). *Hydrology and Earth System Sciences* **14** (1): 1–24 DOI: 10.5194/hess-14-1-2010

Wöhling T, Vrugt JA. 2011. Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. *Water Resources Research* **47** (4) DOI: 10.1029/2010WR009265

Wong JS, Razavi S, Bonsal BR, Wheater HS, Asong ZE. 2017. Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrology and Earth System Sciences* **21** (4): 2163–2185 DOI: 10.5194/hess-21-2163-2017

Wood EF, Roundy JK, Troy TJ, van Beek LPH, Bierkens MFP, Blyth E, de Roo A, Döll P, Ek M, Famiglietti J, *et al.* 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research* **47** (5): W05301 DOI: 10.1029/2010WR010090

Wu Y, Chen J, Wu Y, Chen J. 2012. An operation-based scheme for a multiyear and multipurpose reservoir to enhance macroscale hydrologic models. *Journal of Hydrometeorology* **13** (1): 270–283 DOI: 10.1175/JHM-D-10-05028.1

Xie P.P, Arkin P.A. 1996. Global monthly precipitation: An intercomparison of several datasets based on gauge observations, satellite estimates and model predictions. Eighth Conference on Satellite Meteorology and Oceanography: 225-229.

Yang Z-L, Niu G-Y, Mitchell KE, Chen F, Ek MB, Barlage M, Longuevergne L, Manning K, Niyogi D, Tewari M, *et al.* 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research* **116** (D12): D12110 DOI: 10.1029/2010JD015140

Yassin F, Razavi S, Elshamy M, Davison B, Sapriza-Azuri G, Wheater H. 2019. Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrology and Earth System Sciences* **23** (9): 3735–3764 DOI: 10.5194/hess-23-3735-2019

Yassin F, Razavi S, Wheater H, Sapriza-Azuri G, Davison B, Pietroniro A. 2017. Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. *Hydrological Processes* **31** (19): 3320–3333 DOI: 10.1002/hyp.11267

Yassin F, Razavi S, Wong JS, Pietroniro A, Wheater H. 2019. Hydrologic-Land Surface Modelling of a Complex System under Precipitation Uncertainty: A Case Study of the Saskatchewan

River Basin, Canada. *Hydrology and Earth System Sciences Discussions*: 1–40 DOI: 10.5194/hess-2019-207

Yates D, Sieber J, Purkey D, Huber-Lee A. 2005. WEAP21—A demand-, priority-, and preference-driven water planning model. *Water International* **30** (4): 487–500 DOI: 10.1080/02508060508691893

Yigzaw W, Li H-Y, Demissie Y, Hejazi MI, Leung LR, Voisin N, Payn R. 2018. A new global storage-area-depth dataset for modeling teservoirs in land surface and earth system models. *Water Resources Research* 54: 10,372-10,386 DOI: 10.1029/2017WR022040

Yoon Y, Beighley E. 2015. Simulating streamflow on regulated rivers using characteristic reservoir storage patterns derived from synthetic remote sensing data. *Hydrological Processes* **29** (8): 2014–2026 DOI: 10.1002/hyp.10342

Zaitchik BF, Rodell M, Reichle RH. 2008. Assimilation of GRACE terrestrial water storage data into a land surface model: results for the Mississippi River Basin. *Journal of Hydrometeorology* **9** (3): 535–548 DOI: 10.1175/2007JHM951.1

Zajac Z, Revilla-Romero B, Salamon P, Burek P, Hirpa FA, Beck H. 2017. The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology* **548**: 552–568 DOI: 10.1016/J.JHYDROL.2017.03.022

Zhang K, Kimball JS, Nemani RR, Running SW. 2010. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resources Research* **46** (9) : W09522 DOI: 10.1029/2009WR008800

Zhang S, Gao H, Naz BS. 2014. Monitoring reservoir storage in South Asia from multisatellite remote sensing. *Water Resources Research* **50** (11): 8927–8943 DOI: 10.1002/2014WR015829

Zhang X, Srinivasan R, Arnold J, Izaurralde RC, Bosch D. 2011. Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework. *Hydrological Processes* **25** (14): 2313–2320 DOI: 10.1002/hyp.8058

Zhang X.B, Vincent L.A, Hogg W.D, Niitsoo A. 2000. Temperature and precipitation trends in Canada during the 20th century. Atmos Ocean, 38(3): 395-429

Zhao G, Gao H, Naz BS, Kao S-C, Voisin N. 2016. Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in Water Resources* **98**: 16–31 DOI: 10.1016/j.advwatres.2016.10.014

LIST OF PUBLICATIONS

The following papers were published during candidature:

- Journal Papers:

Yassin, F., Razavi, S., Wheater, H., Azuri G., Davison B. & Pietroniro A. (2017). Enhanced Identification of Hydrologic Models using Streamflow and Satellite Water Storage Data: A Multi-objective Calibration Approach, *Hydrological Process*, DOI: 10.1002/hyp.11267

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G. and Wheater, H. (2019a). Representation of Water Management in Hydrological and Land Surface Models. *Hydrol. Earth Syst. Sci. Discussions.*, 1–35 DOI: 10.5194/hess-2019-7

Yassin, F., Razavi, S., Wong, J. S., Pietroniro, A. & Wheater, H. (2019b). Hydrologic-Land Surface Modelling of a Complex System under Precipitation Uncertainty: A Case Study of the Saskatchewan River Basin, Canada, *Hydrol. Earth Syst. Sci. Discuss.*, 1–40, DOI:10.5194/hess-2019-207

Haghnegahdar, A., Razavi, S., **Yassin, F**. & Wheater, H. (2017). Multi-criteria sensitivity analysis as a diagnostic tool for understanding model behavior and characterizing model uncertainty. *Hydrological Process*. doi:10.1002/hyp.11358

Sheikholeslami, R., **Yassin, F**., Lindenschmidt, K.-E. Razavi, S. (2017). Improved Understanding of River Ice Processes Using Global Sensitivity Analysis Approaches. *J. Hydrol. Eng*., DOI: 10.1061/(ASCE)HE.1943-5584.0001574

Berry, P., **Yassin, F**., Belcher, K. & Lindenschmidt, K.-E. (2017). An Economic Assessment of Local Farm Multi-Purpose Surface Water Retention Systems under Future Climate Uncertainty. *Sustainability* 9(3), 456. Multidisciplinary Digital Publishing Institute. DOI:10.3390/su9030456

Berry, P., **Yassin, F**., Belcher, K. & Lindenschmidt, K.-E. (2017). An economic assessment of local farm multi-purpose surface water retention systems in a Canadian Prairie setting, *Appl. Water Sci.* 1–18. Springer Berlin Heidelberg. DOI:10.1007/s13201-017-0592-7

Berry, P., **Yassin, F**., Grosshans, R. & Lindenschmidt, K.-E. (2017). Surface water retention systems for cattail production as a biofuel. *J. Environ. Manage*. 203, 500–509. DOI:10.1016/j.jenvman.2017.08.019

- Conference Publications/Presentations:

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G. and Wheater, H. Towards an Improved Parameterization of Reservoir Operation for Large-Scale Hydrological Modelling, *AGU Fall Meeting*, Dec 2018.

Yassin, F., Anis, R., Razavi, S., and Wheater, H. Towards an improved representation of reservoirs and water management in a land surface-hydrology model, *AGU Fall Meeting*, Dec 2017.

Yassin, F., Razavi, S. Wheater, H. Improved representation of water management and reservoirs in a land surface-hydrology model, *CGU Meeting*, Jul 2017.

Yassin, F., Razavi, S. Wheater, H. Azuri, G. Davison, B. and Pietroniro, A. 2016. Enhanced Identification of Hydrologic Models using Streamflow and Satellite Water Storage Data: A Multi-objective Calibration Approach, *CGU Meeting*, Jun 2016.

Yassin, F., Razavi, S. Wheater, H. Azuri, G. Davison, B. and Pietroniro, A. 2016. Enhanced Identification of Hydrologic Models using Streamflow and Satellite Water Storage Data: A Multi-objective Calibration Approach, *AGU Fall Meeting*, Dec 2015.

Yassin, F., Wheater, H., Razavi, S., Azuri, G., Davison, B., and Pietroniro, A. Comprehensive, Process-based Identification of Hydrologic Models using Satellite and In-situ Water Storage Data: A Multi-objective Calibration Approach. *EGU General Assembly*, Apr 2015.

Yassin, F., Wheater, H., Azuri, G. Nazemi, A. Davison B. and Pietroniro A. 2014. Towards improved large-scale hydrological modelling application to the Saskatchewan River Basin, *AGU Fall Meeting*, Dec 2014.

Yassin, F., Razavi, S., Wheater, H. Modelling of complex system of Saskatchewan River Basin using Hydrological Land Surface Model. *Changing Cold Regions Network* Annual Modelling Workshops (2014-2018)

PERMISSIONS FOR USE OF PUBLISHED MANUSCRIPTS

Permission for chapter 2

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Jun 04, 2019

This Agreement between Fuad Yassin ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4601820121586
License date	Jun 04, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Hydrological Processes
Licensed Content Title	Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach
Licensed Content Author	Fuad Yassin, Saman Razavi, Howard Wheater, et al
Licensed Content Date	Aug 9, 2017
Licensed Content Volume	31
Licensed Content Issue	19
Licensed Content Pages	14
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	TOWARDS IMPROVED HYDROLOGIC LAND SURFACE MODELING: ENHANCED MODEL IDENTIFICATION AND INTEGRATION OF WATER MANAGEMENT
Expected completion date	Jul 2019
Expected size (number of pages)	200
Requestor Location	Fuad Yassin 812-541 5th ave N Sasktoon, SK s7n5z9 Canada Attn: Fuad Yassin
Publisher Tax ID	EU826007151
Total	0.00 CAD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing

transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Permission for chapter 3 and chapter 4

Licence and copyright agreement

The following licence and copyright agreement is valid for any article submitted from **6 June 2017** onwards. For manuscripts submitted before, the CC BY 3.0 License was used.

Author's certification

By submitting a manuscript, the authors certify that they have read and agreed to the following terms:

- The authors are authorized by their co-authors to enter into these arrangements.
- The work is original and has not been formally published before (except in the form of an abstract, preprint, or as part of a published lecture, review, or thesis), that it is not under consideration for publication elsewhere, that its publication has been approved by all the author(s) and by the responsible authorities – tacitly or explicitly – of the institutes where the work has been carried out, and that the article does not infringe copyright or any other rights by third parties.
- The work does not contain content that is unlawful, abusive, or constitute a breach of contract or of confidence or of commitment given to secrecy.
- The authors warrant that they secure the right to reproduce any material that has already been published or copyrighted elsewhere.
- They agree to the following licence and copyright agreement:

Copyright

- Authors retain the copyright of the article. Regarding copyright transfers please see below.
- Authors grant Copernicus Publications an irrevocable non-exclusive licence to publish the article electronically and in print format and to identify itself as the original publisher.
- Authors grant Copernicus Publications commercial rights to produce hardcopy volumes of the journal for sale to libraries and individuals.
- Authors grant any third party the right to use the article freely as long as its original authors and citation details are identified.
- The article is distributed under the Creative Commons Attribution 4.0 License. Unless otherwise stated, associated published material is distributed under the same licence:

Creative Commons Attribution 4.0 License

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

Under the following conditions:

Attribution — You must give appropriate credit, provide a link to the licence, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licenser endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the licence permits.

Notices:

- The licenser cannot revoke these freedoms as long as you follow the licence terms.
- You do not have to comply with the licence for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.
- No warranties are given. The licence may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

- The CC BY License, of which 4.0 is the recent version, was developed to facilitate open access – namely, free immediate access to, and unrestricted reuse of, original works of all types.
- Under this liberal licence, authors agree to make articles legally available for reuse, without permission or fees, for virtually any purpose. Anyone may copy, distribute, or reuse these articles, as long as the author and original source are properly cited. Thus, CC BY facilitates the dissemination, transfer, and growth of scientific knowledge.
- Please read the full legal code of this licence.

Copyright transfers

Many authors have strict regulations in their contract of employment regarding their works. A transfer of copyright to the institution or company, as well as the reservation of specific usage rights, is typical. Please note that in the case of open-access publications in combination with a Creative Commons License, a transfer of the **copyright** to the institution is possible, as it belongs to the author anyway and is not subject to the publisher.

Any **usage rights** are regulated through the Creative Commons License. As Copernicus Publications uses the Creative Commons Attribution 4.0 License, anyone (the author, their institution/company, the publisher, as well as the public) is free to copy, distribute, transmit, and adapt the work as long as the original author is given credit (see above). Therefore, specific usage rights cannot be reserved by the author or their institution/company, and the publisher cannot include a statement "all rights reserved" in any published paper.

A copyright transfer from the author to their institution/company can be expressed in a special "copyright statement" on the title page of the manuscript, after the abstract and before the introduction. Authors are asked to include the following sentence: "The author's copyright for this publication is transferred to *institution/company*".

Crown copyright

The licence and copyright agreement of Copernicus Publications respects the Crown copyright. For works written by authors affiliated with the British Government and its institutions, a copyright statement will be included at the end of the publication. Authors are asked to use the following statement, which has been approved by the Information Policy department of The National Archives:

> The works published in this journal are distributed under the Creative Commons Attribution 4.0 License. This licence does not affect the Crown copyright work, which is re-usable under the Open Government Licence (OGL). The Creative Commons Attribution 4.0 License and the OGL are interoperable and do not conflict with, reduce or limit each other.
>
> © Crown copyright YEAR

Reproduction request

All articles published by Copernicus Publications have been licensed under the Creative Commons Attribution 4.0 License since 6 June 2017 (see details above) or under its former version 3.0 since 2007. Under these licences the authors retain the copyright. There is no need from the publisher's side to allow/confirm a reproduction. We suggest contacting the authors to inform them about the further usage of the material. In any case, the authors must be given credit.